Colburn on Covert Influences

DONALD W. BRUCKNER
Penn State University, New Kensington

In 'Autonomy and Adaptive Preferences', Ben Colburn claims that preferences formed through covert influences are defective. I show that Colburn's argument fails to establish that anything is wrong with preferences formed in this manner.

INTRODUCTION

In 'Autonomy and Adaptive Preferences', Ben Colburn claims that an adaptive preference is formed as a result of covert influences, and that this is what is wrong with adaptive preferences.[1] I shall argue that Colburn's argument trades on an important ambiguity. Once this ambiguity is exposed, we shall see that his account does not succeed in showing that anything is wrong with adaptive preferences.

I. COVERT INFLUENCE AND COLBURN'S EQUIVOCATION

On Colburn's account, a person's preference is 'covertly influenced when the explanation for [that preference] is something that is necessarily hidden from [him], in the sense that it would *not* be the explanation for [that preference] if it weren't hidden' (p. 64). Colburn gives the example of the formation of a preference to eat some ice-cream as a result of single-frame images of ice-cream being projected during a film one is watching. The cause of the preference for ice-cream is hidden from the agent and necessarily so: if it were not hidden, it would not exert its influence and be the right explanation of the preference (p. 65).

So far so good. The idea is that a preference is covertly influenced if it is formed in response to hidden factors and such that the influence that those factors have is dependent on their being hidden. Thus, Colburn observes, there is a gap between the explanation the agent would give for a preference of hers that was so formed and the explanation that would be given by a third-party observer.

Earlier in his paper, Colburn claims that 'autonomy consists in deciding for oneself what is valuable, and living one's life in accordance

Thanks to Ben Colburn and an anonymous referee for very helpful feedback on an earlier version of this article.

[1] *Utilitas*, 23 (2011), pp. 52–71, at 68. All unadorned page references in the text are to this article.

© Cambridge University Press 2011
Utilitas Vol. 23, No. 4, December 2011
doi:10.1017/S095382081100029X

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press

with that decision' (p. 62). Let us accept that characterization of autonomy for the sake of this argument. Colburn now claims that '[c]overt influence on an agent's [preferences] is bad for autonomy, because it undermines the extent to which we can say she herself is deciding on what is valuable' (p. 66). Consider a pair of examples to make it clear how, on Colburn's view, such an agent is not deciding for herself what is valuable.

In the first example, the agent *is* deciding for herself what is valuable: someone asks a devoted harpist why she is playing that musical instrument and she responds that she devoted herself to it several years ago or simply, 'Because playing the harp well is valuable' (p. 66). What is key here, says Colburn, is that the explanation the harpist gives is the same as the explanation a third party would give for why the harpist is playing the harp. Both the harpist and the third party could give either of the preceding explanations. The explanations from those two perspectives coincide and there are no covert influences.

In the second example, the agent is *not* deciding for herself what is valuable: Someone 'is subliminally influenced into wanting ice-cream' (p. 66). In such a case

> the third-person explanation will disregard her attempts at rationalizing her sudden longing for raspberry ripple. In such cases the 'real' reason for her [preference] is opaque from her first-person point of view. Because something else (about which she cannot know) explains her [preference], we can't say that she is deciding for herself. (p. 66)

Part of this is clear enough. The rationalization the ice-cream desirer would give clearly would not cite the subliminal influences. Yet it is not at all clear why this should matter to whether or not the preference is suspect, and it is here that I think Colburn has committed his equivocation. When asking the agent why she wants the ice-cream, we are not asking for a *causal* reason but for a *justificatory* reason. Indeed, central to his earlier account of deciding for oneself what is valuable is that one would endorse those values upon reflection (p. 63). Endorsing a value would seem to involve pointing out some positive property of the object of value, or some good-making characteristic, such as that the ice-cream is sweet and smooth, or that one would be refreshed by eating a scoop now. Those are justificatory reasons and are the sort of reasons we seek from the agent when asking why, in the justificatory/rationalization sense of 'why', she wants the ice-cream.

On the other hand, in seeking from a third-party observer the 'real' reason for the agent's desire for ice-cream, we seek a different sort of reason altogether. We seek a *causal* reason. And surely Colburn is right that the causal reason given by the third party (the subliminal influence) will be different from the justificatory reason given by the

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press

agent. But it is fallacious to make the inference Colburn explicitly makes from (1) 'something else (about which she cannot know) explains her [preference]' to (2) 'we can't say that she is deciding for herself'. The causal explanation of her preference is a descriptive matter. Whether the agent is deciding on her values for herself is a normative matter, as Colburn's account of deciding for oneself – endorsing those values upon reflection – indicates. Surely it smacks of the genetic fallacy to say that the preference owes its existence to a certain chain of causes and effects (i.e. (1) above) so, therefore, the preference cannot be justificatorily rooted in the values the agent would endorse upon reflection (i.e. (2) above). But it seems that this is just what Colburn has said in the offset passage above: 'Because (1), (2)'.

II. BACK TO COVERT INFLUENCE

We are now in a position to see that when Colburn originally elucidates the idea of a covert influence, he is already trading on two different senses of a reason, explanatory and justificatory. Making out this case will strengthen the reading just given that saddles Colburn with an illicit move from a descriptive premiss to a normative conclusion. Consider what Colburn says about the ice-cream case in initially explaining the idea of a covert influence:

> In such cases, the explanation for their preference for ice-cream is covert: they cannot be aware of it. It seems likely that from their point of view this preference is based on a proper appreciation of the virtues of ice-cream. At any rate, if their preference is genuine they *won't* think that the only reason they have for it is that single-frame images of ice-cream have been interspersed with the film that they have been watching. (pp. 64–5)

This passage contains a subtle but damaging shift from explanatory to justificatory. In the first sentence, Colburn seems to be *explaining* how the preference comes about. In order for it to come about as it does, the person who is subliminally influenced cannot, of course, be aware of the influence. The second sentence admits of two readings. On the first, 'based on' means 'is explained by': it seems likely that from the agent's point of view the preference *is explained by*, that is, caused by, the agent's appreciation of the virtues of ice-cream. So on this reading, the first and second sentences are both addressed to causal *explanation*. On the second reading, however, 'based on' means 'is justified by': it seems likely that from the agent's point of view the preference *is justified by* the agent's appreciation of the virtues of ice-cream. The third sentence, however, seems clearly to admit only of a *justificatory* interpretation. In that sentence, Colburn addresses 'the only reason they have for' the preference. Having a reason *for* a preference is to be able to *justify* it. If

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press

Colburn had meant the reason in an *explanatory* sense, he would have had to leave off the 'for' and addressed 'the only [causal] reason they have' the preference. So it is the equivocation in the second sentence between explanatory and justificatory senses of 'based on' that allows Colburn to move from the explanatory claim in the first sentence to the claim about justification in the third sentence.

So, Colburn's slide from explanatory reasons to justificatory reasons when he initially introduces the concept of a covert influence makes it likely that later, when he applies that concept to show a particular agent (the one I discussed in Section I) is not deciding for herself (and hence not autonomous), he makes the same equivocation there.

Colburn might respond that I am wrong to claim that the third sentence admits only a justificatory reading, and that we can read the passage as explanatory throughout.[2] The locution 'reason for' is not necessarily tied to justification. For example, the expression indicates an explanatory reason in 'the only reason for the moon's moving as it does is gravity'. So it seems plausible that the third sentence is meant in an explanatory mood, as are the first two.

Such a response, however, only creates a new problem. In talking about whether a 'preference is genuine' in the third sentence, I take it that on Colburn's view we are talking about a normative matter. Again, the preference is genuine, or it is *hers* in a robust sense, or the agent is deciding 'for herself' what is valuable, only if the agent would endorse it upon reflection. But endorsing the preference upon reflection is a normative matter, not a descriptive one. One does not offer the right sort of endorsement if one merely recognizes as descriptively correct a certain causal explanation for one's preference. That would not be deciding for oneself what is valuable. Rather, the agent endorses the preference if 'she judges (or would judge) of some such things that they are valuable' (p. 63). And that is a normative judgement.

So, the problem created if we insist on an explanatory reading of the third sentence is that it is opposed to the normative claim about genuineness that Colburn wants to make in that sentence. Indeed, in the next section, I go on to show that an agent might very well have a preference that is genuine while recognizing that 'the only [causal] reason they have for it is that single-frame images of ice-cream have been interspersed with the film that they have been watching', contrary to what Colburn claims on an explanatory reading of the sentence I have been discussing.

[2] Indeed, Colburn has offered such a response in a very careful and constructive referee report on an earlier version of my article.

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press

III. COVERT INFLUENCE AND AUTONOMY

Finally, let us see how the operation of a covert influence need not undermine autonomy as Colburn claims. Again, let us take as given Colburn's claim that 'autonomy consists in deciding for oneself what is valuable, and living one's life in accordance with that decision'. As I mentioned above, Colburn requires, in order for one to be deciding for oneself what is valuable, that one would endorse one's values upon reflection. As well, 'her reflection is, or would be if it took place, free from factors which limit the extent to which we can say that she is deciding for herself' (p. 63). Colburn's brief gloss of this second (independence) condition indicates that the reflection must be free from such factors as hypnotic suggestion, subliminal influences, or coercive persuasion, and he introduces the concept of covert influence in order to provide a necessary condition for independence, namely, lack of covert influence (p. 64). I want to contend that the preference for ice-cream formed in response to subliminal messaging *can* be fully autonomous on his account.

Suppose we have an agent who has watched a film with those single-screen images of ice-cream and this has resulted in a preference to eat some ice-cream. As she is about to place her ice-cream order, we inform her of the genesis of her desire and warn her that (Colburn suspects) she may not be acting autonomously. So we ask her whether she really endorses the desire she has for ice-cream, now that she knows how it came about. She informs us that she does endorse that desire because acting on it would be refreshing to her. She sticks to this claim even when we press her on the fact that she only *has* this desire as a result of the subliminal messages. Be that as it may, she says, she still finds the ice-cream attractive because it will refresh her.

She seems to have passed the first of Colburn's hurdles to count as deciding for herself what is valuable, for she endorses her value upon reflection. We now need to check whether 'her reflection is ... free from factors which limit the extent to which we can say that she is deciding for herself.' At first glance, it seems clearly not, since subliminal influence was involved and this is one of the things that Colburn explicitly says is a defeating factor. Yet when asking whether any of those defeating factors apply, we are asking whether any of them apply to the circumstances of her *reflection* on her values (a justificatory matter), not the circumstances of the *formation* of her values (a causal matter). Clearly the *formation* of her value was subject to subliminal influence. But we may easily enough imagine that her *reflection* is not subject to any of those defeating factors. So she decides upon the value for herself. The genesis of the value was outside her control, but she

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press

acts autonomously in pursuing this value given that she endorses it upon reflection.[3]

Perhaps what Colburn needs to add to his conception of a covert influence is a repudiation condition. Recall that in his official characterization of a covert influence, the causal influence must be hidden from the agent, for revelation of the influence would render the influence causally non-efficacious. But I have shown that an influence can be covert in *this* sense yet not affect the agent's autonomy, as long as she endorses the resulting preference upon reflection, even in light of revelation of the causal genesis of the preference. So perhaps Colburn should want to say that such an influence is not damaging to autonomy. Rather, he should call an influence 'covert' only if, in addition to its actual operation being necessarily hidden from the agent, the agent would repudiate the preference resulting from the influence, were the agent informed of that influence after the preference was formed in response to that influence.[4] It seems to me that this solution might be workable. Pursuing it would require careful reformulation of much of Colburn's original discussion in order to avoid the equivocation I claim to have exposed in that discussion, and require, as well, trials against test cases of the sort of adaptive preferences that philosophers have found worrisome.[5]

[3] Colburn considers such a case as I have discussed in the text, in which the real causal reasons for the formation of the preference for ice-cream are revealed to the agent, but she retains it anyway (p. 65). It seems to me that his discussion there may also trade on an ambiguity between causal and justificatory reasons. As well, further complications arise in interpreting what he says there, complications arising from a distinction he ignores between the causal explanation for the *genesis* of a preference and the causal explanation for the *retention* of a preference. Nothing material to the major elements of my critique hinges on how we interpret Colburn's analysis of this case, so I refrain from discussing the interpretative possibilities in the text.

[4] In Colburn's feedback on my critique and on p. 65 of his paper, there are indications that he is sympathetic to the addition of this repudiation condition.

[5] An anonymous referee points out that adaptive preference formation – in the broadest sense of altering one's preferences in light of available options – need not be restricted to covert influences, i.e. that there can exist adaptive preferences not formed by covert influences. So once we have restricted adaptive preferences to covert influences, both Colburn and I have left the adaptive preferences debate behind. It does seem that Colburn intended to say that all adaptive preferences are due to covert influences (p. 52), but perhaps we can interpret him most charitably to say that of all preferences formed in the light of available options ('adaptive preferences' in the referee's broad sense), the ones that have something wrong with them are the ones that are due to covert influences. So we can read Colburn as trying to provide an analysis of what is wrong with a certain subset of adaptive preferences in the referee's broad sense. Thus, it is not clear that the adaptive preference debate has been left behind. Indeed, see the Conclusion for further tie back.

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press

CONCLUSION

Even granting Colburn's claim that the formation of adaptive preferences is a paradigm case of covert influence (as he originally formulates the notion), if I am right that his discussion is infused with a (clearly unintentional but nevertheless damaging) equivocation, then he has failed to show that there is anything wrong with adaptive preferences. Nevertheless, I believe that he has isolated something important to autonomy in the requirement that an agent endorse her values if she reflects (or were she to reflect) on them. Indeed, this is exactly why I think there is not anything wrong with adaptive preferences that meet just that endorsement condition.[6]

dwb12@psu.edu

[6] I have argued this more fully in my 'In Defense of Adaptive Preferences', *Philosophical Studies* 142 (2009), pp. 307–24. If I were to revise that discussion in light of Colburn's paper and my critique, I would be more explicit that the reflection must be on both the *content* of the preference and the *process* by which it was formed.

https://doi.org/10.1017/S095382081100029X Published online by Cambridge University Press