

Defusing the Miners Paradox

Michael Shaffer

Abstract

This paper presents a case for the claim that the infamous miners paradox is not a paradox. This contention is based on some important observations about the nature of ignorance with respect to both disjunctions and conditional obligations and their modal features. The gist of the argument is that given the uncertainty about the location of the miners in the story and the nature of obligations, the apparent obligation to block either mine shaft is cancelled.

1. Introduction

In this paper a more nuanced and accurate construal of the miners paradox is presented and on this basis the Miners paradox is defused. This involves understanding some important points about rational obligation, disjunction and uncertainty. The main contentions made here are based on the observation that crucial modal and epistemic dimensions of the story are totally absent in typical presentations of the paradox. Specifically, these modal and epistemic dimensions are left out of the typical formalizations of the disjunctive knowledge involved and the conditional obligations that are at the heart of the alleged paradox. When these notions are included in the formal translation of the story and when they are added in it turns out that there is no paradox in the miners' story at all. This manner of dissolving the miners paradox is to be preferred to alternative solutions—particularly that of Kolodny and MacFarlane (2010)—on the basis of its relative simplicity. Importantly, it does not require radical revisions of the logical of indicative conditionals and the rejection of the unrestricted validity of *modus ponens*, as Kolodny and MacFarlane's solution requires. Let us begin then by focusing on some important aspects of knowledge as they pertain to disjunction.

2. Disjunctions and Uncertainty

Consider the following story:

Joe wakes up the day after the 2000 U.S. presidential election. He has not followed the details of the race and knows only that there were two candidates being voted for: George W. Bush and Al Gore. He is aware that only one of them could have won, he does not know which one won and has no evidence to favor either the claim that Bush won or the claim that Gore won. Joe meets up with his buddy Tony and Tony asks Joe “Who won the election?” Joe responds with “Either Bush or Gore.”

In light of this brief story, consider the following parsing of Joe’s assertion, where we understand clearly that the component sentences involved are contingent:

(BG) Either Bush won the 2000 U.S. presidential election or Gore won the 2000 U.S. presidential election.

This ordinary language English sentence might be understood to have the following richer correlate:

(BGA) Either Bush *actually* won the 2000 U.S. presidential election or Gore *actually* won the 2000 U.S. presidential election.

More formally, where W stands for “Bush won the 2000 U.S. presidential election” and G stands for “Gore won the 2000 U.S. presidential election,” we can regiment BGA simply as follows, where “■” is an actuality operator and “ \vee ” is standard disjunction:

(BGA1) ■W \vee ■G

Additionally, and as explicitly noted in the Bush/Gore story, it is clear that the sort of disjunction involved does not permit it to be that case that Bush won and Gore won, so we need to amend things as follows based on the recognition that $\neg\Diamond(W \ \& \ G)$, where “&” is standard conjunction and “ \neg ” is standard negation:

Defusing the Miners Paradox

$$(BGA2) (\blacksquare W \vee \blacksquare G) \& \neg(\blacksquare W \& \blacksquare G).$$

In ordinary discourse the use of sentences like BGA also seems to connote epistemic uncertainty with respect to the truth of the disjuncts involved.¹ If this is the case, then with respect to Joe, BGA can be regimented as follows, where K_jx is “Joe knows that x ”:

$$(BGA3) (\blacksquare W \vee \blacksquare G) \& \neg(\blacksquare W \& \blacksquare G) \& (\neg K_j \blacksquare W \& \neg K_j \blacksquare G).$$

This is just the claim that one of either Bush or Gore actually won the election but Joe does not know which of Bush or Gore actually won the election. Finally, there is an implicative connotation involved that it is possible that Bush won and that it is possible that Gore won and that the utterer knows this. The disjunction concerning the actuality of Bush winning or the actuality of Gore winning is not supposed to be true in virtue of the fact that one of the disjuncts is impossible and the other actually true. So the epistemically complex, modalized, use of disjunction in such contexts suggests the following rather complex rendering, where “ \diamond ” is the orthodox possibility operator of modal logic:

$$(BGA4) (\diamond W \& \diamond G) \& (K_j \diamond W \& K_j \diamond G) \& (\blacksquare W \vee \blacksquare G) \& \neg(\blacksquare W \& \blacksquare G) \& (\neg K_j \blacksquare W \& \neg K_j \blacksquare G).²$$

This appears to be a typical rendering of the epistemically and modally rich use of “or” in cases involving contingent statements where there is uncertainty and this has important implications for the Miners paradox, which has received much attention of late in the context of both ethics and epistemology.³ What it suggests is that the ordinary language usage of disjunction involves important epistemic and modal content that is overlooked in the standard logical translations of disjunction.

¹ This is clearly the case when the epistemic agent in question does not know the disjunction to be true on the basis of knowing that one of the disjuncts is true but not knowing anything about the truth value of the other or does not know the disjunction to be true on the basis of knowing one disjunct to be true and employing weakening by disjunction introduction. The view of disjunction developed here is then closely related to the modal account of disjunction developed by Zimmerman (2000) and Geurts (2005) and inspired by Kamp (1973).

² Again, this view is then closely related to that defended in Zimmerman (2000) and in Geurts (2005).

³ See Parfit (manuscript), Kolodny and MacFarlane (2010) and Dutant and Fitelson (manuscript).

3. The Miners Paradox

Here is a version of the story that gives rise to the miners paradox. It is essentially the same as that presented, for example, by Kolodny and MacFarlane (2010):

Ten miners are trapped in a flooding mine; they are either all in shaft A or all in shaft B. Given Tony's information, each location is equally likely. Tony has just enough sandbags to block one shaft. If the miners are in the blocked shaft, they will all be saved. If the miners are in the other shaft, then they will all be killed. If Tony does nothing, the water will distribute between the two shafts, killing only the one miner at the lowest level of the mine.⁴

The paradox implicit in this sort of story is supposed to be derived from considerations raised initially by Jackson (1991) about possible cases where the alternative with the best outcome does not have the highest utility on the body of known information in the situation so described. Let us then turn to the presentation of the alleged paradox.

On the basis of this story the following claims seem to be true:

- (M1) Tony ought to block neither shaft.
- (M2) If the miners are in A, Tony ought to block A
- (M3) If the miners are in B, Tony ought to block B.
- (M4) Either the miners are in A or they are in B.

(M2)–(M4) entail,

- (M5) Either Tony ought to block A or Tony ought to block B.

Prima facie, the paradox and the disjunctive uncertainty involved in the miners story can be formally regimented as follows, where O_{Tx} is "Tony is rationally obligated to do x",⁵ M_A is "miners are in A", M_B is "miners are in B", B_A is "block shaft A", B_B is "Block shaft B and " \rightarrow " is standard implication:

⁴ See also Parfit (manuscript), Regan (1980) and Pettersson (2014).

⁵ "Rational obligation" is just meant here to indicate some rationally mandated action that follows from one or more rational principles.

- (M*1) $O_{T \neg} B_A \ \& \ O_{T \neg} B_B$.
 (M*2) $M_A \rightarrow O_T B_A$.
 (M*3) $M_B \rightarrow O_T B_B$.
 (M*4) $M_A \vee M_B$.
 (M*5) $O_T B_A \vee O_T B_B$.
 (M*6) $\neg \diamond (B_A \ \& \ B_B)$.
 (M*7) $\neg (\blacksquare M_A \ \& \ \blacksquare M_B)$.
 (M*8) $\diamond B_A \ \& \ \diamond B_B$.⁶

So, there appear to be conflicting obligations in this case. Technically, this is not a contradiction. Generating a contradiction from M*1–M*5 requires additional steps as follows:

- (M*9) $O_{T \neg} B_A \rightarrow \neg O_T B_A$.
 (M*10) $O_{T \neg} B_B \rightarrow \neg O_T B_B$.

M*1, M*9 and M*10 imply this:

- (M*11) $\neg O_T B_A \ \& \ \neg O_T B_B$.

Given these additional steps the miners paradox then at least appears to be a bona fide paradox.

4. Solution

However, even this appearance is deceptive in light of the epistemic modalities and uncertainties involved in the use of “or” in the story and in the parsing of the conditional obligations that are crucial to the story. This comports with the point made above about the ordinary language implicature associated with disjunction. When we incorporate these facts into the regimentation of the paradox we get the following, more complex characterization of the miners paradox propositions:

- (M*1) $O_{T \neg} B_A \ \& \ O_{T \neg} B_B$.
 (M*2') $(\blacksquare M_A \ \& \ K_T M_A) \rightarrow O_T B_A$.
 (M*3') $(\blacksquare M_B \ \& \ K_T M_B) \rightarrow O_T B_B$.

⁶ M*6, M*7 and M*8 are not required for deriving the miners paradox, but, given the account of disjunctive uncertainty proposed here, they are parts of the story. Moreover, they are crucial parts of the solution to the alleged paradox of the miners.

- (M*4') ($\blacksquare M_A \vee \blacksquare M_B$) & ($\neg K_T M_A \ \& \ \neg K_T M_B$) & ($K_T \diamond M_A \ \& \ K_T \diamond M_B$).
 (M*5) $O_T B_A \vee O_T B_B$.
 (M*6) $\neg \diamond (B_A \ \& \ B_B)$.
 (M*7) $\neg (\blacksquare M_A \ \& \ \blacksquare M_B)$.
 (M*8) $\diamond B_A \ \& \ \diamond B_B$.⁷

Notice here that the important changes are to be found in M*2', M*3' and M*4' and they importantly involve discrimination of what is actually the case and what is known to be the case in the story. The conjunction of M*4', M*7 and M*8 is analogous to BGA4 and it reflects the same kind of context involving the use of disjunction in light of epistemic modalities and uncertainty we found in the Bush/Gore story. As such, it is appropriate to make these changes in the same way. As per the miners' story then, the agent involved in the situation does not have information favoring either $\blacksquare M_A$ or $\blacksquare M_B$ and so does not know M_A and does not know M_B , although he knows both M_A and M_B are possible and that just one must be actually true.⁸ Thus, M*4' captures better the position of that agent with respect to the disjunction involved in the miners story and his/her knowledge with respect to the disjuncts.

Crucially then, the next contention made here is that in M*2' and M*3' the rational obligation to block one but not both of the mine shafts is conditionally dependent on the agent's *knowing that* M_A or M_B , respectively.⁹ In other words they are subjective obligations. This bit of absolutely fundamental information is absent in the initial presentation of the putative paradox. It is, however, an entirely plausible and principled assumption. If the miners are in shaft A or in shaft B respectively, but Tony does not know this, then Tony cannot reasonably be thought to have an obligation to block that shaft in question rather than the other. To deny this principle in general would impose a plethora of unknown and/or unknowable obligations on every agent, the resultant objective obligations would be practically worthless in deliberations about what to do under conditions of limited information and it would involve violations of a plausible version of the "ought implies can" principle. This latter point follows because such obligations would be obligations that the agent *could not*—in the sense of epistemic

⁷ The putative paradox can also be presented in terms of justified belief rather than knowledge, but this changes nothing about the analysis of the case and about the solution proffered to it here.

⁸ Moreover, he has not reasons to favor the truth of M_A over M_B or M_B over M_A .

⁹ See Fischer and Ravizza (1998), Ginet (2000), Rosen (2008) and Mele (2011) for philosophical defenses of this claim.

possibility—meet in light of this sort of ignorance. They would be obligations that are epistemically impossible to meet in the sense that they are required of the agent even though the agent is totally unaware of them. Simply consider the following scenario involving such a conditional obligation:

Bill is taking a walk in the woods. He is near a river, but his view of the river is totally obscured. It is also loud. So, he cannot see or hear anything in the river. Sally, who cannot swim, has fallen in the river near enough to Joe so that he could physically reach her and easily save her. He is an excellent swimmer.

If obligations like those that are alleged to pertain to the miners in the simple version of the putative paradox are legitimate, then by parity of reasoning we would have to say that Bill is obligated to save Sally where he could easily save her, but where he has no knowledge that she is drowning. But it is clear that his epistemic state defeats the conditionality of that obligation and it is not even remotely plausible to claim that he does, in fact, have such an obligation. He is exculpated from that obligation in virtue of his ignorance (which is no fault of his) and this is an utterly typical but reasonable sort of excuse in such cases. Specifically, the conditional obligation is defeated by such ignorance and such agents rightly can claim that they had no such obligation when the agent is unaware of it and is not at fault in being unaware of it.¹⁰ The same thing then goes for the miners paradox conditional obligations. As such, where $O_S R \neg p$ is “S should bring it about that $\neg p$ ” and p is a factual state with negative consequences such that it is morally bad that p , the position defended here is that obligations with the general form $(p \ \& \ \neg K_S p) \rightarrow O_S R \neg p$ are not (at least not always) obligations and that reasonable obligations (at least typically) have the form $(p \ \& \ K_S p) \rightarrow O_S R \neg p$.¹¹ But, notice then that given the epistemic uncertainty involved in M^*4' and this more accurate rendering of the obligations involved, the inference from M^*2' , M^*3' and M^*4' to M^*5 is invalid. One cannot derive a rational, subjective, obligation to block A or to block B from the fact that the miners *might be in A* and *might be in B* when the agent does not know either

¹⁰ This contention also has strong empirical support as Kissinger-Knox, Aragon and Mizrahi (2018) demonstrates.

¹¹ See Spencer and Wells (forthcoming) for discussion and defense of rational obligations and knowledge requirements like the one suggested here.

possibility to be the case and does not have any reason(s) to favor the truth of M_A over M_B or of M_B over M_A .¹² This is the very kind of ignorance that defeats the obligations involved. So, there is no paradox here and given M^*1 , M^*2' , M^*3' and M^*4' it is clear that, rationally, Tony ought not to block shaft A and ought not to block shaft B in light of his ignorance about the location of the miners.

The last matter that needs to be addressed here concerns Kolodny and MacFarlane's (2010, 118–119) rather convoluted contention that we cannot properly interpret the conditional obligations involved in the miners paradox as subjective obligations in the manner suggested here. They do so on the basis of the claim that subjectivist interpretations of conditional obligations "...cannot make good sense of the use of "ought in advice" (Kolodny and MacFarlane 2010, 119)." That is to say that if such obligations were subjective in the sense suggested here, we would not be able to make sense of advice about situations where advice is given by agents in superior information states to agents in inferior information states with respect to the very same situation. Consider their Dialogue 1, where the agent in the miners paradox has an exchange with an adviser who, *ex hypothesi*, knows the location of the miners:

AGENT: I ought to leave both shafts open, guaranteeing that nine survive.

ADVISER: No, you ought to block shaft A. Doing so will save all ten of the miners (Kolodny and MacFarlane 2012, 119).

Kolodny and MacFarlane contend that the subjective construal of conditional obligations cannot properly make sense of this sort of exchange for the following reasons.

They claim, first, that the subjectivist construal of conditional obligations makes sense of Agent's assertion in Dialogue 1 given his/her limited information. But, second, they contend that this is not true of Adviser's assertion in Dialogue 1. With respect to this scenario, they rightly claim that Adviser is not making a claim about what Agent ought to do *given Agent's limited information*, for Agent already knows that and then Agent and Adviser would not be in disagreement when Adviser challenges Agent's assertion. In order to make sense of this, Kolodny and MacFarlane claim that

¹² Notice that the conditional obligations also do not follow where the undesirable state is known merely to be possible.

the advice from Adviser only makes sense if there is real disagreement between Agent and Adviser. Adviser knows where the miners are located and so challenges Agent's claim to be obligated not to block either shaft and, according to Kolodny and MacFarlane, this makes sense only if Adviser is really in disagreement with Agent. After claiming that this desideratum cannot be met, they argue further that the subjectivist reading cannot be saved by claiming that Agent acquires evidence about the location of the miners upon hearing Adviser's claim and so is no longer obligated to not block the shafts due to Agent's ignorance of the location of the miners. But, this latter argument is really a bit of a red herring, for it is easy to make sense of the disagreement between Agent and Adviser in terms of the subjectivist construal of conditional obligations.

First, let us refer to Agent as "Tony" and Adviser as "Vivian". So, on the subjectivist reading, in Dialogue 1 Tony has the following obligations: $O_{T \rightarrow B_A} \& O_{T \rightarrow B_B}$. This is only the case, however, because of his ignorance of the location of the miners as we saw previously and regimented as follows: $(\blacksquare M_A \vee \blacksquare M_B) \& (\neg K_T M_A \& \neg K_T M_B) \& (K_T \diamond M_A \& K_T \diamond M_B)$. The crucial subjectivist bases for the conclusion that Tony ought to block neither shaft are the following claims: $(\blacksquare M_A \& K_T M_A) \rightarrow O_{T B_A}$ and $(\blacksquare M_B \& K_T M_B) \rightarrow O_{T B_B}$. However, Vivian's situation is entirely different. From the perspective of her information state the following claims are true: $O_{V \rightarrow B_A}$, $K_V M_A$, $\blacksquare M_A$ and $(\blacksquare M_A \& K_V M_A) \rightarrow O_{V B_A}$. Vivian's obligation would then be entirely different from Tony's if she were in a position to act to save the miners and knows what she knows. But, Tony does not know the location of the miners and so does not have the same conditional obligation as Vivian. Kolodny and MacFarlane appear to reject subjectivism, at least in part, on the basis of the following utterly implausible claim: $(\blacksquare M_A \& K_V M_A) \rightarrow O_{T B_A}$. More importantly, they contend on this basis that the subjectivist *cannot* explain the sense of disagreement in Dialogue 1. But this is simply not true. The disagreement between Tony and Vivian is easy to understand in terms of subjectivism about conditional obligations, independent of any worries about how Vivian's assertion effects Tony's evidential state.

Vivian disagrees with Tony about what Tony should do, but only in the sense that *Tony's obligations* can be understood relative to two possible information states that *Tony* could be in and only one of which he is actually in. So, Vivian is simply disagreeing with Tony in the sense that she is saying something like this: "No. You should block shaft A, because *if you knew what I know* that would be your correct obligation." But, Tony's information

state is such that he does not know what Vivian knows. If he in fact learns what Vivian knows, then his correct conditional obligation would change with the alteration in his information state and it would (at least in normal cases) conform to the subjective obligation that Vivian herself has given her information state (i.e. to block shaft A). So, they disagree about what is the right thing to do only in the sense that they derive different subjective obligations for Tony, but they do so on the basis of the different information states Tony could be in. What Vivian is saying is simply that Tony could be in a better state of information (one that she, in fact, occupies) and if that were the case, then Tony would no longer be obligated to block neither shaft. Of course, he is not in that state though in Dialogue 1 and since he isn't in that information state he does not have the obligation to block only shaft A. He is exculpated from the obligation to do that because of his impoverished information state, and this would not be the case if he were in Vivian's information state. Vivian's advice then is nothing more than the specification of an epistemic possibility that is not currently an actuality for Tony and her advice is nothing more than an entreaty to Tony to improve his information state. So, this objection does not really undermine the subjectivist interpretation of conditional obligations. The sense of disagreement involved in Dialogue 1 is simple to understand and given the reasons in favor of the subjectivist interpretation of conditional obligations discussed earlier, the solution to the miners paradox presented here is to be preferred to alternatives that are far less conservative.

References

- Dutant, J. and Fitelson, B. (manuscript). Knowledge Centered Epistemic Utility Theory. <http://fitelson.org/boulder.pdf>.
- Fischer, J. and Ravizza, M. (1998). *Responsibility and Control*. Cambridge: Cambridge University Press.
- Geurts, B. (2005). Entertaining Alternatives: Disjunctions as Modals. *Natural Language Semantics* 13: 383–410.
- Ginet, C. (2000). The Epistemic Requirements for Moral Responsibility. *Nous* 34: 267–277.
- Jackson, F. (1991). Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics* 101: 461–82.
- Kamp, H. (1973). Free Choice Permission. *Proceedings of the Aristotelian Society* 74: 57–74.

Defusing the Miners Paradox

- Kissinger-Knox, A, Aragon, P. and Mizrahi, M. (2018). Does Non-Moral Ignorance Exculpate? Situational Awareness and Attributions of Blame and Forgiveness. *Acta Analytica* 33: 161–179.
- Kolodny, N. and MacFarlane, J. (2010). Ifs and Oughts. *Journal of Philosophy* 107: 115–43.
- Mele, A. (2011). Moral Responsibility for Actions: Epistemic and Freedom Conditions. *Philosophical Explorations* 13: 101–111.
- Parfit, D. (manuscript). What We Together Do.
http://individual.utoronto.ca/stafforini/parfit/parfit_what_we_together_do.pdf.
- Pettersson, K. (2014). Informational Models in Deontic Logic: A Comment on “Ifs and Oughts” by Kolodny and MacFarlane. *Filosofiska Notiser* 1: 91–103.
- Regan, D. (1980). *Utilitarianism and Cooperation*. New York: Oxford University Press.
- Rosen, G. (2008). Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility. *Journal of Philosophy* 105: 591–610.
- Spencer, J. and Wells, I. (forthcoming). Why Take Both Boxes? *Philosophy and Phenomenological Research*.
- Zimmermann, T. (2000). Free Choice Disjunction and Epistemic Possibility. *Natural Language Semantics* 8: 255–290.

Michael Shaffer
Department of Philosophy
CH 365
St. Cloud State University
720 4th Ave. South
St. Cloud, MN 56301
shaffermphil@hotmail.com