It Can Be Irrational to Knowingly Choose the Best

J. DMITRI GALLOW *

In some decisions, the verdicts of orthodox causal decision theory (CDT) are *unstable* in the sense that they change depending upon how likely you think you are to choose each of the available options. Sometimes, this means that CDT's permissions are fickle. CDT says that A is permitted; but as soon as you become confident you'll choose A, CDT retracts the permission and says that A is forbidden. Other times, CDT's forbiddances are fickle. It says that A is forbidden, but as soon as you become confident you'll choose A, CDT retracts the forbiddance and says that A is permitted.

Several have objected to CDT's instability and proposed a stable alternative called 'MaxRat'.[1] Jack Spencer thinks MaxRat is *too* stable. He thinks that, while a decision theory should not be fickle with its permissions, it *should* be fickle with its forbiddances. Spencer argues that we should reject MaxRat because it conflicts with the principle

Knowingly If you know that you will choose an option, x, and you know x is better than every other option available to you, then it is permissible for you to choose x.

I agree with Spencer that defenders of MaxRat should reject this principle. However, I disagree insofar as he thinks that he and orthodox causalists are in a better position to accept it. Both orthodox CDT and his own heterodox theory also contradict **Knowingly**.

1. *Instability*

CDT says that you should choose whichever option has the highest *utility*, U, where the utility of an option, x, is

$$U(x) \overset{\text{def}}{=} \sum_k C(k) \cdot D(x \wedge k).$$

Draft of September 7, 2022; Word Count: 2,250

* Thanks to Jack Spencer for helpful feedback.

1. See Barnett, 2022, Gallow, 2020, and Podgorski, 2022.

Here, each k is a state of nature, C is your credence function, and D measures your desires. If you're in the state k, then CDT says that the objective value of choosing x is given by $D(x \wedge k)$, so $U(x)$ is your subjective expectation of x's objective value.

Consider the following decision.

> Amelia must choose between two envelopes labelled 'a' and 'b'. Yesterday, her fairy godfather made a prediction about which envelope she would choose. If he predicted a ('k_a'), then he put \$1 in a. If he predicted b ('k_b'), then he put \$2 in b. Amelia is certain that he predicted correctly.

Conditional on Amelia choosing a, she knows for sure that her fairy godfather predicted she would choose a. And, conditional on her choosing b, she knows for sure that he predicted she would choose b. Then, her credence in k_a will be her credence that she chooses a, and her credence in k_b will be her credence that she chooses b,

$$C(k_a) = C(a) \quad \text{and} \quad C(k_b) = C(b).$$

Assuming Amelia's desires are linear in dollars, the utility of a is her credence that she takes a, and the utility of b is two times her credence that she takes b,

$$U(a) = C(a) \quad \text{and} \quad U(b) = 2 \cdot C(b).$$

Whether $U(a)$ or $U(b)$ is higher depends upon Amelia's opinions about what she will choose. So long as she is confident enough that she'll select a/b, CDT says a/b is required.

In a choice between two options, x and y, Barnett, Gallow, and Podgorski all advise you to look at a family of *conditional* utilities, $U_y(x)$, where

$$U_y(x) \stackrel{\text{def}}{=} \sum_k C(k \mid y) \cdot D(x \wedge k)$$

is the utility x has, conditional on you selecting y. Then, $R(x, y) \stackrel{\text{def}}{=} U_x(x) - U_x(y)$ measures how much better than y you will expect x to be, conditional on you choosing x. ('R' for *r*atifiability.) If $R(x, y) > R(y, x)$, then these authors say that x is required. Spencer names this rule '**MaxRat**'. Applied to Amelia's decision, $R(a, b) = U_a(a) - U_a(b) = 1$ and $R(b, a) = U_b(b) - U_b(a) = 2$, so **MaxRat** says that b is required, no matter which option Amelia thinks she'll choose.

In contrast, Spencer (2021) says that, in a choice between two options, if you currently expect that x is better than y, and you would *continue* to think this as you grow more confident that you will select x, then x is required. That is, if both $U(x) > U(y)$ and $U_x(x) > U_x(y)$, then x is required. Call this '**Spencer's Rule**'. Applied to Amelia's decision, if she is 80% sure that she'll choose a, then $U(a) = 0.8 > 0.4 = U(b)$ and $U_a(a) = 1 > 0 = U_a(b)$. So **Spencer's Rule** says that a is required.

2. *Spencer's Argument Against MaxRat*

Suppose Amelia *knows* that she will take a. Because she knows that the predictions are accurate, she knows that she will take a iff a contains $1 and b contains $0. So, if she knows that she will take a, then she knows that a is better than b. So **Knowingly** says that it is permissible to take a. But **MaxRat** disagrees. So **MaxRat** and **Knowingly** are inconsistent. But Spencer contends that **Knowingly** is "undeniable". So Spencer concludes that we should reject **MaxRat**.

Someone might object that it's not possible for Amelia to know that she will take a, but I don't think we should. A resolute intention to choose a gives Amelia fantastic evidence that she will choose a. She may have a long history of always following through on her resolute intentions. Barring inductive scepticism, Amelia is in a position to know that she'll choose a. Nor should it matter whether this choice is rational. In general, we should acknowledge that it is possible to know that you'll choose x, even if x is irrational. Some may object that it's not possible for Amelia to know that the prediction is accurate, but I don't think we should. So long as her choices are predictable, there's no reason why Amelia herself couldn't come to know that they've been accurately predicted. We could grant all this but deny that Amelia can know that a is better than b. But, again, I do not think we should. We should acknowledge that knowledge is closed under competent deduction. Suppose that Amelia formed her belief that a is better than b by competently deducing it from her knowledge that she will take a and her knowledge that the prediction is accurate, together with her knowledge that, if she will take a and the prediction is accurate, then a is better than b. Then, Amelia should know that a is better than b. Even if we have worries about the closure of knowledge in general, there shouldn't be concerns about its application to this particular case.

Knowingly then tells us that it is permissible for Amelia to take a, which conflicts with **MaxRat**. I do not wish to reject any of these assumptions, so I accept the conclusion: **MaxRat** contradicts **Knowingly**.

3. *An Argument Against CDT and Spencer's Rule*

But **MaxRat** is not alone. With similar assumptions, both **Spencer's Rule** and orthodox CDT also contradict **Knowingly**. Consider the following decision:

> Casey must choose between two envelopes labelled 'c' and 'd'. Yesterday, his fairy godmother made a prediction about which he would choose. If she predicted c, ('k_c'), she put $1 in c and nothing in d. If she predicted d ('k_d'), she put $\$1/\epsilon$ (for some $\epsilon > 0$) in d and nothing in c.

Suppose Casey's credence that he will take c is $1 - \epsilon$, and he knows that he will take c. Assuming that Casey knows that the predictions are accurate, he knows he will take c iff c contains $1 and d contains nothing. So, if he knows he will take c, then he knows that c contains more money than d, and so he knows that c is better than d. So **Knowingly** says that it is permissible for Casey to take c.

Both orthodox CDT and **Spencer's Rule** disagree. Because Casey's credence that he will take c is $1-\epsilon$, his credence in k_c is $1-\epsilon$, and his credence in k_d is ϵ. So the utility of c is

$$\begin{aligned} U(c) &= C(k_c) \cdot D(c \wedge k_c) + C(k_d) \cdot D(c \wedge k_d) \\ &= (1-\epsilon) \cdot 1 + \epsilon \cdot 0 \\ &= 1-\epsilon \end{aligned}$$

whereas the utility of d is

$$\begin{aligned} U(d) &= C(k_c) \cdot D(d \wedge k_c) + C(k_d) \cdot D(d \wedge k_d) \\ &= (1-\epsilon) \cdot 0 + \epsilon \cdot (1/\epsilon) \\ &= 1 \end{aligned}$$

So orthodox CDT says that c is impermissible, and d is required. Since the utility of d would *continue* to exceed the utility of c if Casey selects d, $U_d(d) = 1/\epsilon > 0 = U_d(c)$, **Spencer's Rule** agrees. So both contradict **Knowingly**.

You may object that it's not possible for Casey to know that he will take c. I disagree. I think we are typically in a position to know that we will choose an option on the basis of our intention to do so, independent of whether that option is rational. And I do not think that this knowledge requires certainty; for small enough ϵ, having a credence of $1-\epsilon$ that you will choose x does not preclude you from knowing that you'll choose x, so long as your evidence is good enough, and you in fact *do* choose x. But suppose you disagree, and you think that Casey cannot know he will take c. Just for illustration, suppose you think that knowledge is sensitive to the stakes of your practical situation. Then, like Fantl & McGrath (2002), you may say that you know ϕ only if it is rational for you to act as if ϕ. And you may suggest that, since it is *not* rational for Casey to act as if he will take c, he cannot know that he will take c. Let's not dwell on the fact that Fantl & McGrath restrict their principle so that it does not apply when you exercise causal control over whether ϕ. The important point is this: a defender of **MaxRat** could just as plausibly say that it is not rational for *Amelia* to act as if she will take a. Applying Fantl & McGrath's principle in the same way, they could then conclude that Amelia cannot know she will take a. Sauce good for the goose is good for the gander. If pragmatic encroachment gives Spencer plausible grounds to deny that Casey knows what he'll choose, it gives a defender of MaxRat equally plausible grounds to deny that Amelia knows what she'll choose.

Some may object that it's not possible for Casey to know that the prediction is accurate. Again, I don't think we should. But, more importantly: it looks like any plausible reason you may have for denying that Casey can know the prediction made about *him* is accurate is a reason a proponent of **MaxRat** could use to deny that Amelia can know that the prediction made about *her* is accurate.

We could grant both that Casey knows that he will take c and that he knows that the prediction is accurate. Even so, we could deny that Casey is in a position to know that

c is better than d, by denying that knowledge is closed under competent deduction. But I don't see any plausible reason to deny *this* instance of closure which couldn't equally well be offered as a reason to deny the instance of closure we used in arguing that Amelia was in a position to know that a is better than b.

So it seems to me that the argument that orthodox CDT and **Spencer's Rule** contradict **Knowingly** is just as strong as Spencer's argument that **MaxRat** contradicts **Knowingly**. And it seems to me that the defences available to orthodox CDT and **Spencer's Rule** are equally available to **MaxRat**.

4. *Knowledge and Expectation*

Spencer (forthcoming) does not discuss decisions like Casey's, but he does offer a reason to think that Casey cannot know that c is better than d. When discussing an unrelated objection to **Knowingly**, Spencer claims that you cannot know that x is better than y if $U(y) > U(x)$ (fn. 16). I believe the idea is this: utility is your subjective expectation of objective value. So, if $U(y) > U(x)$, then y is better than x in expectation. And if y is better than x in expectation, then you cannot know that x is better than y.

Knowledge and expectation If your expectation of y's value is greater than your expectation of x's value, then you cannot know that x is better than y.

Compare: if your expectation of Emilia's height is greater than your expectation of Grant's height, then you cannot know that Grant is taller than Emilia. This principle says that Casey cannot know that c is better than d, for $U(d) > U(c)$. And it allows that Amelia can know that a is better than b, since $U(a) > U(b)$.

We should reject **Knowledge and expectation**. The principle can sound more plausible than it should if we're not careful to distinguish what's expected in the everyday sense from what's expected in the sense of mathematical expectation. Suppose you know for sure that Grant is 6 feet tall, and you're $1 - \epsilon$ sure that Emilia is 5 feet tall, but there's an ϵ probability that she's been zapped by a growth ray gun and is now $2/\epsilon$ feet tall. Then, your expectation of Emilia's height is greater than your expectation of Grant's height. But you do not, in the everyday sense, *expect* Emilia to be taller than Grant—after all, you're nearly *certain* that Grant is taller than Emilia! Moreover, so long as you can know that the ϵ probability event did not obtain, you can know that Grant is taller than Emilia. And in exactly the same way, you can know that x is better than y, in spite of the fact that your expectation of y's value is greater than your expectation of x's.

5. *Knowing*

Spencer (forthcoming, fn 17) briefly discusses the principle

Knowing If you know that x is better than every other option available to you, then it is permissible for you to choose x.

Spencer's theory of rational choice contradicts **Knowing.** To appreciate why, consider this decision:

> Imogen must choose between two envelopes, labelled 'i' and 'j'. If her fairy godfather predicted she'd choose i, he put \$10 in j and nothing in i. If he predicted j, he put \$2 in i and \$1 in j. Imogen is certain he predicted correctly.

In Imogen's decision, $U_i(i) < U_i(j)$ and $U_j(j) < U_j(i)$. So **Spencer's Rule** is silent. But Spencer's full theory of rational choice is not. It tells Imogen she must choose j. Roughly, in a choice between two options, x and y, if $U_x(x) < U_x(y)$ and $U_y(y) < U_y(x)$, then Spencer says: if $U_x(x) > U_y(y)$, then x is required (see Spencer, 2021, for details and caveats). In Imogen's decision, $U_i(i) = 0 < 1 = U_j(j)$. So Spencer's theory tells Imogen that she is required to take j.

But now suppose Imogen knows that she will choose j. Then, she knows that her fairy godfather predicted that she'd choose j, and so knows that i contains \$2 whereas j contains only \$1. So she knows that i is better than every other option available to her. **Knowing** says that it is permissible for Imogen to choose i. But Spencer disagrees. So he rejects **Knowing.** In his fn 17, Spencer says that decisions like Imogen's are counterexamples to **Knowing.**

Speaking for myself: I feel the intuitive pull of **Knowing** and **Knowingly** equally. And I'm just as inclined to take Amelia's decision to be a counterexample to **Knowingly** as I am inclined to take Imogen's to be a counterexample to **Knowing.** If it is a cost of a theory that it rejects **Knowingly,** it seems to me that it should likewise be a cost that a theory rejects **Knowing.** Moreover, once **Knowing** is denied, it seems to me that there's little additional cost to denying **Knowingly** as well.

References

Barnett, David James. 2022. "Graded Ratifiability." In *The Journal of Philosophy*, **119** (8): 57–88. [1], [2]

Fantl, Jeremy & McGrath, Matthew. 2002. "Evidence, Pragmatics, and Justification." In *The Philosophical Review*, **111** (1): 67–94. [4]

Gallow, J. Dmitri. 2020. "The Causal Decision Theorist's Guide to Managing the News." In *The Journal of Philosophy*, **117** (3): 117–149. [1], [2]

Podgorski, Abelard. 2022. "Tournament Decision Theory." In *Noûs*, **56** (1): 176–203. [1], [2]

Spencer, Jack. 2021. "Rational Monism and Rational Pluralism." In *Philosophical Studies*, **178**: 1769–1800. [2], [6]

Spencer, Jack. forthcoming. "Can It Be Irrational to Knowingly Choose the Best?" In *Australasian Journal of Philosophy*. [5], [6]