A Higher-Order Approach to Diachronic Continence

Catherine Rioux
University of Toronto

Abstract: We often form intentions to resist anticipated future temptations. But when confronted with the temptations our resolutions were designed to withstand, we tend to revise our previous evaluative judgments and conclude that we should now succumb—only to then revert to our initial evaluations, once temptation has subsided. Some evaluative judgments made under the sway of temptation are mistaken. But not all of them are. When the belief that one should now succumb is a proper response to relevant considerations that have newly emerged, can acting in line with one’s previous intention nonetheless be practically rational? To answer this question, I draw on recent debates on the nature of higher-order evidence and on what rationally responding to such evidence involves. I propose that agents facing temptation often have evidence of “deliberative unreliability”, which they ought to heed even when it is “misleading” (that is, even when their evaluative judgments are in fact proper responses to the relevant considerations then available). Because evidence of deliberative unreliability can “dispossess” agents of normative reasons for evaluative judgments and actions that they would otherwise have, being continent despite judging that one should now succumb can often be more rational than giving in.

Key words: intention, weakness of will, higher-order evidence, practical rationality, decision, epistemic rationality

We often form intentions in anticipation of future temptations: think of those who resolve to go for an early morning run despite knowing the attractions of a warm bed, or of the dieter who decides to abstain from eating desert before entering a restaurant. In a cool hour, we conclude that we ought to refrain from temptation and form intentions accordingly. But when placed in “tempting situations” or confronted with the objects of our desires, we often experience temporary shifts in our evaluative judgments: despite initially judging that it is “all-things-considered” best to drink only two espressos in the morning instead of four (especially in light of financial and health reasons) and resolving accordingly, I come to revise my evaluative judgment after having had my second cup of coffee, only to later regret my coffee decisions. Or consider the case of Liam, an agent who has resolved not to drink at his friend’s wedding because he plans to run a marathon two days later and wants to be in good shape. Liam knows that his evaluative judgment will shift once at the wedding, and that he will think that he should start drinking. However, he also knows that after the wedding his judgment will shift back, and he will regret having drunk if he did. Suppose at \( t_1 \) (before attending the wedding) Liam sees his future act of succumbing at \( t_2 \) (once at the wedding) as certain. In that case, it seems that Liam should avoid the wedding altogether: by not attending at all, he won’t ruin his chances of running the marathon. From Liam’s perspective, however, this should appear to be an unfortunate conclusion: it seems that Liam would do better by his own lights if he had the capacity, at \( t_2 \), not to act in accord with his temporary preferences. At the same
time, if Liam acts in line with his previous resolution, he would then go against his present judgment and thus be akratic.

Liam and similarly placed agents appear to be caught in a double bind: their previous resolution seems to impact what they ought to do in the moment of temptation, but their following through on their earlier intention seems to amount to akrasia—and thus to a form of practical irrationality. In fact, agents in Liam’s position face what has come to be known as a “temptation problem”:\(^1\) we would like to explain how their acting in line with their earlier resolutions can count as practically rational and well supported by their normative reasons, despite their judgment that they ought to succumb also being in phase with the balance of relevant considerations—especially those having newly emerged at the time of temptation. As we shall see, on a picture on which rationality (both practical and epistemic) is a matter of one’s reactions being properly responsive to one’s possessed normative reasons,\(^2\) agents in Liam’s position not only ought to act in line with their previous resolutions, but should also ultimately revise their evaluative judgment that they ought to succumb to temptation. Still, as I shall argue, agents who hold on to evaluative judgments formed under the influence of temptation will display a greater capacity to respond to their possessed normative reasons if they act contrary to these judgments, and in line with their prior resolutions, than if they give in.

To argue for this conclusion, I shall draw on resources from contemporary epistemology, and in particular on recent debates surrounding evidence of one’s own unreliability.\(^3\) I propose that, in the moment of temptation, agents in Liam’s practical situation often possess a specific kind of “higher-order evidence”: namely, evidence that their capacity to correctly evaluate reasons for action is impaired. I argue that rational agents ought to respond to that evidence, act in line with their initial evaluations, and ultimately revise the evaluative judgments that were made under the influence of temptation, even when these judgments are in fact appropriate responses to the balance of relevant and then accessible normative considerations. Attending to the role of higher-order evidence of deliberative unreliability highlights previously underappreciated connections between epistemic and practical rationality, due to higher-order evidence’s capacity to dispossess agents of reasons that they would otherwise have, for both evaluative beliefs and actions.

1. EVALUATIVE JUDGMENT UNDER TEMPTATION’S INFLUENCE

One might think that temptation cases such as Liam’s are in fact less puzzling than they first appear: if agents confronted with temptation simply end up with mistaken views about what is “all-things-considered” best to do, they may not be practically irrational in going against their current judgments. Cases such as Liam’s might actually be thought of as instances of “inverse akrasia” (Arpaly 2003), in which agents’ motivation manages to track relevant normative considerations that they overlook in their judgments. On this way of construing Liam’s case and similar ones, the practical situation of agents when they act in accord with their resolutions is analogous to the situation of Huckleberry Finn (see Arpaly 2003: 75–79), who refuses to turn Jim in despite believing that it is the right thing to do. Since both Liam and Huck fail to appreciate the full force of relevant, epistemically accessible practical considerations, they are led astray in their normative beliefs. We do not need to appeal to evidence of one’s own “deliberative unreliability” to see why going against such misguided evaluative judgments can constitute a response to one’s possessed normative reasons for action.

I agree that some cases of rational continence in the context of temporary judgment shift seem to share the structure of “inverse akrasia” cases. However, not all instances of the previously described “temptation problem” seem to fit nicely into that category. When precisified enough, many cases that share the central features of Liam’s example seem to involve a correspondence between the agent’s evaluative judgment at the time of temptation and relevant, epistemically ac-
cessible considerations. As Richard Holton (2009: chap. 7) emphasizes, being placed in a “situation of temptation” can cause agents to revise their evaluative judgments in ways that are not necessarily irrational. This is because acquaintance with temptation can both provide genuinely new information and reinforce one’s already existing desires for the tempting option: when faced with the object of one’s desires, one can realize that the reward of succumbing is greater than anticipated, or that the good for which one was holding out is not as valuable as it first appeared. Thus, on a plausible construal of Liam’s case and similar ones, acquaintance with temptation shifts the balance of relevant, epistemically accessible considerations regarding what the agent should do. Admittedly, from Liam’s perspective, there still exists a benefit to persisting in his resolution: by refraining from drinking, Liam can improve his chances of running his personal best. Only, that benefit is now outweighed by more important considerations to which Liam has gained access in the moment of temptation (such as how great celebrating with his friends would be).

With this picture of the most puzzling temptation cases in clear view, we can restate our main guiding question: can sticking to one’s previous resolution be more rational than giving in, even when one would thereby be acting against a current evaluative judgment which is in fact properly grounded in relevant and then epistemically accessible considerations? By drawing on an account of the impact of “misleading” self-doubt on rational agents’ judgments and actions, I shall argue for a positive answer to this question. Even when one’s current evaluative judgment that one should succumb is a proper response to relevant, accessible facts (and one’s evidence of deliberative unreliability is thus “misleading”), one can be rationally required to act in line with one’s previous intention. To be sure, agents who follow through on their previous intention without revising their current judgment that they should succumb are less than fully rational. Still, they are better off, from the point of view of rationality as responsiveness to reasons, than if they were to simply succumb to temptation. As we shall see, this is due to the “dispossessing power” of misleading evidence of deliberative unreliability, which can wedge a gap between relevant facts to which agents have epistemic access and the actual normative reasons they possess.

2. NORMS ON INTENTION STABILITY VERSUS EVIDENCE OF DELIBERATIVE UNRELIABILITY

Before developing my account of practically rational continence in cases of temporary judgment shift, I want to briefly explore how a competing, influential view would handle our cases. At first blush, a promising way of vindicating the rationality of continence in our temptation cases might be to endorse the existence of diachronic norms on “intention stability”, such as the following norm defended by Michael Bratman (2012: 79):

\[(D) \text{ The following is locally irrational: Intending at } t_1 \text{ to } X \text{ at } t_2; \text{ throughout } t_1 - t_2 \text{ confidently taking one’s relevant grounds to support this very intention adequately; and yet at } t_2 \text{ newly abandoning this intention to } X \text{ at } t_2.\]

\(D\) is supposed to be a wide-scope, local constraint on rationality. Since the past is fixed, one cannot at \(t_2\) satisfy (\(D\)) by changing the fact that at \(t_1\) one intended to \(X\) at \(t_2\). However, one can satisfy (\(D\)) either by sticking to one’s prior intention or by revising one’s assessment of one’s grounds. It might seem that since Liam does not fulfill the antecedent of the conditional, he is in fact “off the hook” regarding \(D\): as we just saw, on a plausible construal of his case, when in the moment of temptation, Liam does not confidently take his relevant grounds to support his previous resolution. However, to salvage the rationality of continence in our cases, we might be drawn to supplement \(D\) in the spirit of Bratman’s view, namely by attributing to the agent what Bratman calls a “concern for self-governance” (2012: 77–84; 2014: 302–303). As Bratman defines this concern, agents who exemplify it care that their behavior be guided by a unified diachronic practical standpoint, consti-
tuated by their intentions and policies. By attributing such a concern to the agent, we might manage to portray her as fulfilling D’s antecedent after all: if the end of self-governance provides the agent with further reasons to act on her past intention (in addition to the reasons she considered in her initial deliberation), then even in the moment of temptation, she might still in fact take her relevant grounds to support her past intention.

I believe that we should be unsatisfied with the strategy just sketched. Giving oneself extra reasons for continence by adopting the end of self-governance seems akin to a form of “distal self-management” (Ferrero 2010). Because adopting the end of self-governance impacts the agent’s payoff structure (by raising the cost of succumbing to temptation, understood as a loss in terms of self-governance), agents who maintain their resolve via such an end seem to be introducing “features extraneous to the merits of the case” (Ferrero 2010: 3)—considerations they did not have in mind when they first settled on a course of action. Full diachronic autonomy seems incompatible with stacking the deck against one’s succumbing self in this manner. Instead of relying on a desire for resoluteness that resembles a fetish, it seems preferable to follow through on one’s resolution in response to the reasons initially considered in one’s deliberation. As we shall see, my higher-order evidence proposal aims to do justice to this idea.

I propose that agents in cases such as Liam’s will typically have access to evidence of “deliberative unreliability.” I take such evidence to constitute a species of “higher-order evidence”, defined as evidence questioning the reliability of one’s reasoning abilities or of one’s assessment of the evidence (Christensen 2007, 2010). I hold that agents who display temporary judgment shift will often be in an epistemic situation closely resembling the following:

**Mathematics:** Julia engages in mathematical deductive reasoning to determine how much money she owes her friend Elena. It seems to Julia that she owes Elena $60. However, Julia knows that she has taken a drug that is very likely to make her mathematical abilities unreliable, without her being in a position to notice whether this is so. As it happens, the drug has not actually been effective: Julia’s mathematical reasoning abilities were left intact. The conclusion that she owes Elena $60 is best supported by the relevant facts to which she has access.

I propose that just like Julia, agents in Liam’s type of practical situation will often have evidence of unreliability with respect to their capacity to assess facts bearing on the question they aim to settle. As an athlete seriously committed to a difficult pursuit who aims to abstain from alcohol, Liam knows that his ability to properly respond to his practical reasons and arrive at sound evaluative judgments might very well be corrupted by the presence of temptation: Liam knows that while at social events preceding races, committed marathoners like himself often reach mistaken practical conclusions about how much alcohol they should drink and later experience regret. In short, Liam knows that he belongs to a vulnerable category with respect to good practical deliberation. Unfortunately, however, he is not in a position to know whether he is actually among the unlucky marathoners whose judgment is clouded by temptation. As it happens (and as we saw in section 1), from his perspective at the time of the wedding, succumbing is in fact best supported by the relevant facts to which he has epistemic access, and the evaluative judgment that he should succumb is thus “well-informed”.

I believe that we can offer a similar treatment of many of the most puzzling cases of temporary judgment shift: we can show that they feature agents who possess what has come to be known as “misleading” higher-order evidence. In the set of cases I have in mind, the agent’s evidence of deliberative unreliability can properly be deemed “misleading” because (contrary to what goes on in “inverse akrasia” cases) her evaluative judgment is not out of touch with the normative considerations then available. By drawing on a recent account of how evidence of unreliability—even when misleading—should be factored in by epistemically and practically rational agents, we can explain
why agents in Liam's position are both under an epistemic obligation to revert to their original evaluative judgments and practically rational in their resolve. If Liam based all of his reactions (both evaluative beliefs and actions) on his possessed normative reasons, he would not judge that he ought to succumb. But if he holds on to such a belief anyway, then considering that it is out of kilter with the set of reasons he actually possesses, he does better—from the point of view of rationality as a matter of responding to one's reasons—to act against it than in line with it.

Many hold that misleading higher-order evidence can make it impermissible to react in ways that are nonetheless best supported by the relevant "first-order" considerations. However, there is substantial debate as to how such evidence can manage to defeat one's first-order reasons. After all, evidence of one's unreliability does not seem to act as a "rebutting defeater" (Pollock 1986): in Mathematics, the fact that Julia has taken the drug (which constitutes her higher-order evidence) does not bear on the truth or falsity of the proposition that she owes Elena $60. Nor does evidence of one's unreliability seem to constitute an "undercutting defeater" (Pollock 1986), reducing the weight or strength of the agent's reasons: even considering her higher-order evidence, the first-order, normative facts available to Julia still seem to support her original answer in the strongest possible way, namely by deductively entailing that conclusion. In fact, reflection on the relation between higher-order evidence and coherence requirements on attitudes has led some to doubt the very possibility of higher-order defeat: if agents in Julia's epistemic position both ought to entertain a certain first-order belief (since it is supported by the relevant first-order considerations) and believe (due to their misleading higher-order evidence) that that belief is epistemically irrational, then aren't they doomed to fall short of some rational ideal? My account of how misleading higher-order evidence of deliberative unreliability can rationalize going back to one's original all-things-considered judgment and acting in line with it engages with those issues.

3. FROM DISPOSSESSING DEFEAT TO DIACHRONIC CONTINENCE

To elucidate our puzzling cases, we must determine how misleading higher-order evidence can deprive agents of normative reasons for evaluative judgments and actions that they would otherwise possess. I suggest that cases such as Liam's feature a gap between one's reasons for evaluative judgments and actions, on the one hand, and relevant, epistemically accessible facts, on the other. To fully appreciate this gap, we need to better understand what constitutes a rational response to misleading higher-order evidence.

My view of how rational agents ought to respond to misleading higher-order evidence is rooted in an influential conception of rationality that I alluded to earlier, on which rationality (both epistemic and practical) is a matter of responding to one's "possessed reasons" (Kiesewetter 2017, Lord 2018). On this picture of rationality, facts that count as normative reasons bearing on the rational merits of one's reactions must somehow lie within one's ken: only by standing in a "privileged epistemic relation" to the subject can a fact affect the rationality of her reactions. In addition to this "epistemic condition" on reasons possession, I follow Kurt Sylvan (2015), Errol Lord (2018), and others in accepting the existence of a further, "practical condition" on reasons possession. I hold that to possess R as a reason to X, the agent also has to be in a position to suitably appreciate the favoring relation between the reason-constituting fact R and the relevant reaction X. According to this line of thought, if one does not comprehend how a certain relevant, epistemically accessible fact is connected to a certain relevant reaction (in other words, if one does not see the practical upshot of the relevant fact), then that fact cannot affect the rational merits of one's reactions. To illustrate: even if a layperson meets the epistemic condition regarding the premises of a sophisticated mathematical deduction (for instance, even if she knows that these premises hold), if she cannot competently treat those premises as reasons for the sophisticated deduction's conclusion, then she does not satisfy the practical condition on reasons possession with respect to the premises. Because
the layperson does not possess the premises as reasons to endorse the deduction's conclusion, we cannot view her failure to accept the conclusion as a rational defect.\textsuperscript{13}

I suggest that something similar is happening in Liam's case: misleading higher-order evidence of the sort Liam possesses prevents him from meeting the practical condition on reasons possession with respect to the considerations supporting succumbing that he has newly accessed at the time of temptation. Due to misleading higher-order evidence's "dispossessing power", if—when in the presence of temptation—Liam responds to the set of reasons he actually possesses, then he will in fact respond to the reasons he considered in his original deliberation, before encountering temptation. In short, considering Liam's "uncluttered" set of original reasons, resolve is the rational option.

I hold that the competence, or "know-how", to properly recognize the favoring relation between a fact $R$ and a particular reaction $X$ involves a reliable disposition to treat $R$-like considerations as reasons for $X$-like reactions just in case they are such reasons (see Sylvan 2015, Lord 2018). Taking inspiration from González Prado (2020), I propose that when one has higher-order evidence of "deliberative malfunction", one is unable to manifest the kind of reliable disposition necessary to possess as reasons the considerations that have newly emerged in the presence of temptation. My proposal is grounded in an evaluative perspective centered on dispositions, according to which the goodness of a disposition depends on whether its manifestation leads to "success" in a range of relevant counterfactual cases (see Lasonen-Aarnio 2020, 2021). From that perspective, consider someone who, despite having evidence that her capacity to appreciate a certain kind of support relation is malfunctioning, nonetheless continues to take certain considerations to support her relevant reactions to the same degree as she would were she to have no evidence of unreliability. We should see such an agent as failing to manifest a "good", reliable disposition to appreciate the favoring relation in question. For in a class of relevant counterfactual cases (namely those where the agent's higher-order evidence is not misleading, and her capacity to appreciate the favoring relation is truly impaired), such a disposition for "obstinance" (Lasonen-Aarnio 2020) would lead her to fail with respect to the aim of treating $R$-like considerations as reasons just in case they actually are reasons. To borrow Lasonen-Aarnio's (2020: 625) image, an agent who manifests a disposition for obstinance is like a lifeguard with a disposition not to jump into the pool when it appears that someone is drowning: even if these dispositions can yield "success" in certain particular cases (namely, when appearances are misleading), they are generally unreliable with respect to the achievement of their related aims. As limited creatures, it is not feasible for us to be disposed to disregard our evidence only when it is misleading. This is why the manifestation of a disposition not to heed one's evidence of deliberative unreliability should be seen as a general epistemic defect, of the sort that prevents agents from meeting the competence condition on reasons possession regarding the considerations that have emerged at the time of temptation.

We could also insist (as González Prado [2020] does), that the practical condition on reasons possession involves a "precautionary dimension". The thought is that full competence, both epistemic and of other kinds, involves a higher-order competence to assess one's first-order competence and its required conditions, to ensure that performances are not produced in circumstances where success is too unlikely (see Sosa 2015: chap. 3). Consider a fully competent archer who, given bad lighting conditions, takes herself to be outside of her "reliability range", and so decides not to take a shot at a particular target. Just like this archer, an agent who is fully competent with respect to appreciating the favoring relation would also, when in possession of evidence of deliberative unreliability, deem the conditions too unfavorable for the production of her performance of taking the considerations that are newly accessed in the presence of temptation to support her evaluative judgment. By preventing agents from being in a position to manifest a suitable competence to treat the considerations that are newly accessed in the moment of temptation as reasons, evidence of deliberative malfunction manages to rob agents of reasons that they would otherwise have. Con-
A Higher-Order Approach to Diachronic Continence

tinence and retreat to Liam’s initial evaluative judgment are rational responses to the reasons that he actually possesses: once his higher-order evidence is factored in, Liam is left with the reasons he considered in his initial deliberation, before encountering temptation.

On the present picture of how higher-order evidence can defeat, misleading higher-order evidence does not generate a conflict between the rational ideal of coherence and that of reasons-responsiveness: because higher-order evidence hampers one’s very possession of certain first-order considerations as reasons, retreat to a more cautious first-order attitude (in Liam’s case, to his previous evaluative judgment) is supported both by one’s possessed first-order reasons and by one’s higher-order evidence. Finally, the present view leaves room for the observation that agents won’t always possess evidence of deliberative unreliability when confronted with the temptation they resolved to withstand: just like the young man who, as a pre-adolescent, formed the resolution never to be susceptible to the charms of women (Gauthier 1997: 17), agents will sometimes justifiably consider themselves to be in a better deliberative position at the time of temptation than when they first deliberated. My view can explain why agents in such circumstances are under no rational pressure to follow through. Nonetheless, as I hope to have shown, as long as agents do possess undefeated evidence—even if misleading—of current deliberative unreliability, their resolve and retreat to their initial evaluations will remain practically and epistemically rational. Evidence of one’s unreliability thus affects both rational belief and rational action.

NOTES

1. See Thoma 2018 for a summary of discussions of these cases in the dynamic choice literature.
2. See Lord 2018: chap. 2 for a compelling recent defense of this conception. See also section 3.
3. For a helpful review of these recent debates, see Whiting 2020. For influential explorations of the notion of evidence of one’s own unreliability, see Feldman 2005, Christensen 2010, and Lasonen-Aarnio 2014.
5. Bratman’s thought is that one’s plan-states frequently have cross-temporal contents, referring to related past and future elements of one’s overall plans. When these states play out over time “in the proper way”, the agent is said to govern herself not just at a time but over time. See Bratman 2012: sect. 6).
6. To borrow Holton’s (2009: 118) expression. Holton shares my worries about the type of strategy just outlined.
7. This case is inspired by similar examples discussed (among others) by Christensen (2010), Elga (2013), Lasonen-Aarnio (2014), and González Prado (2020: 324). Discussions of higher-order evidence usually focus on cases where the agent gains evidence of unreliability when already in possession of “first-order” evidence bearing directly on the question at hand. However, as Christensen (2016) points out (relying on a case structurally similar to Mathematics), higher-order evidence that is possessed prior to first-order evidence can also plausibly impact the rationality of belief.
9. For convincing explanations of why defeat by higher-order evidence differs from undercutting and rebutting defeat, see Christensen 2010 and Lasonen-Aarnio 2014.
10. See, in particular, Lasonen-Aarnio 2014.
11. As Lord (2018: chap. 3) notes, the acceptance of an “epistemic filter” on reasons is widespread.
13. See Lord 2018: 97–102 for discussion and further cases supporting the thesis that the epistemic condition is necessary but not sufficient for reasons possession.
REFERENCES


