Against Stochastic Independence of Ethical Theory Valuations

Thomas J. Fournier

Which normative theory should guide our actions if we don't know which one is true? Even with perfect knowledge of the non-normative facts regarding a given choice, we are often left to choose on the basis of normative theories we are uncertain about – theories that even renowned experts disagree over. On many occasions we might hold some credence in various competing ethical theories, yet be required to render a decision and take action. That is, we are forced to accept a theory for the purpose of guiding action, though we are not fully convinced of its truth.

In "Rejecting Ethical Deflationism," Jacob Ross wrestles with this problem and proposes a heuristic for accepting or rejecting ethical theories in regard to sets of options under consideration.[1] He provides a more-detailed explication of that heuristic in his dissertation, *Acceptance and Practical Reason*,[2] in which he uses formal proofs and prose to identify four key principles for theory selection. His dissertation then weaves those four principles into a general formula for the expected value of accepting a given theory under a given set of options. Ross admits that the resulting formula is "rather complicated" for his

[1] Jacob Ross, "Rejecting Ethical Deflationism", *Ethics* 116 (July 2006): 742-768.

[2] Jacob Ross, *Acceptance and Practical Reason*, (Ann Arbor: ProQuest Information and Learning Company, 2007)

stated goal of practical reasoning.[3] Therefore, he goes on to offer a simplification that assumes stochastic independence of the values assigned by competing theories. In other words, he assumes that the expected values assigned to options by a given theory are, in no way, influenced by the expected values assigned by competing theories. This assumption is pivotal in rendering Ross's heuristic practical for action guidance, and its consequences are the source of my theses for this paper.

While I am sympathetic with Ross's quest to develop a theory-acceptance heuristic, I assert that his simplifying assumption is unjustified because it allows acceptance of certain prescriptions given by credence-worthy, non-fanatical theories that, nonetheless, violate the taboos of other credence-worthy, non-fanatical theories. This has repugnant and destabilizing social consequences – an outcome that ethical theory ought to eschew until such (unlikely) time as we have found the one true ethical theory. In raising my concerns about the simplifying assumption of *stochastic independence*, I will uncover two additional shortcomings of Ross's approach. The first has to do with *limiting the domain of the probability principle* to theories – rather than also considering probability of theory prescriptions. The second has to do with *impoverished effectiveness under bi-elemental theory sets* – i.e. when agents only have credence in two theories. The nature of my concerns regarding stochastic independence, probability principle domain and bi-elemental theory sets will be explained in greater detail below; however, prior to considering my arguments, let's overview Ross's well-reasoned principles of theory selection, namely:

[3] Ross, *Acceptance and Practical Reason,* 61.

probability, typicality, selectiveness and ***dispersion***. Due to space considerations, I will provide only the briefest schematic of each without going into their justification.

Under Ross's ***probability principle***, we have pro tanto reason to accept theories we regard as more probably true than theories we regard as less probably true.[4] This acceptance is for practical, action-guiding purposes in the face of uncertainty about theory truth. It is *not* a declaration of theory truth. Also, the reason is called 'pro tanto' in that it is good only as far as it goes under the qualification 'all other considerations being equal'. This qualification is crucial because the other three acceptance principles will not usually yield equal results from theory to theory for a given set of options.

Ross's ***typicality principle*** holds we have pro tanto reason to accept a theory that values options more typically like the valuations of other theories in which we also have credence – as opposed to accepting a theory that values options less typically.[5] This is conceptually understood as throwing out the outliers. Imagine you are trying to decide on a course of action from among a number of options and you hold some level of credence in five different theories. You apply each theory to the options available, and each theory assigns a value to each option. If one theory evaluates options much higher or lower as compared to the other four theories, it may be ruled out. In statistical terms, the typicality is the correlation between the values assigned by a given theory and the values assigned by that theory's complement – whereby a complement is the total group of theories in which

[4] Ross, *Acceptance and Practical Reason*, 45.
[5] Ross, *Acceptance and Practical Reason*, 45.

one has credence minus the given theory being evaluated for typicality.[6] Thus, a theory that does not correlate well with its complement is **a**typical and its prescriptions may be rejected in regard to the particular set of options being considered. Notice that typicality is sensitive to the set of options under consideration. Therefore a theory found acceptable for adjudicating one option set (i.e. decision scenario) might be unacceptable when adjudicating a different option set.

The ***selectiveness principle*** holds that we have pro tanto reason to accept a theory that is more selective than one that is less so – whereby 'selective' means valuing most options within the considered group as being relatively far away from the top of the theory's valuation range. In other words, a selective theory will find that the good options are rare.[7] All other things being equal, selective theories are to be preferred over unselective ones. Ross uses the statistical standardized maximum as a measure of theory selectiveness. This maximum is the measure of the difference between the mean of the values the theory assigns to the options under consideration and the maximum value it assigns to the highest valued option – expressed in units of standard deviation.[8] For future reference in what follows, make a mental note that expressing a theory's selectiveness in units of standard deviations means first finding the difference between the maximum value and the mean of the values, and then dividing that difference by the standard deviation of all the values.

---

[6] Ross, *Acceptance and Practical Reason*, 48. (Note: If we call the group of all theories in which we have credence G and we are evaluating theory A which is a member of G, then A's complement is {G – A}.

[7] Ross, *Acceptance and Practical Reason,* 49, 51. (Note: Ross has a confusing typo on pg. 49 whereby he reverses 'selective' and 'unselective'. This creates an apparent inconsistency within the first full paragraph of pg. 49 in regard to the definition of selective. His further treatment of selective on pg.51 makes clear he meant to label *selective* theories as ones that value most options under consideration as relatively *far* from the top of the theory's valuation range not *near* the top.)

[8] Ross, *Acceptance and Practical Reason*, 51.

Think of this technique as somewhat analogous to dividing a foot by twelve to convert it to units of inches.

Under Ross's ***dispersion principle***, we have pro tanto reason to accept a theory under which the expected values it assigns to the options within the group of options under consideration are widely divergent.[9] In statistical terms, the dispersion of a theory is the standard deviation of the expected values assigned by that theory to the options under consideration. A higher standard deviation implies the theory is better able to detect differences between options as opposed to evaluating them as roughly similar.

Now, as one might expect, the incorporation of *probability, typicality, selectiveness* and *dispersion* into a single expression of expected value for a given theory is a complex task. Ross has done an admirable job of it and derived an all-inclusive theorem with an unfortunate level of *tremendous* complexity.[10] In fact, he's left with a complex summation ranging over every credence-worthy theory applied to every option under consideration and requiring summations of repetitive calculations of means, maximums, standard deviations and correlations. Ross recognizes the impractical nature of such a calculation for our daily deliberations, and he sets about the task of reducing complexity by making a crucial (I think mistaken) assumption, namely: the values assigned to our options by any one theory in which we have some credence are independent of and unaffected by the values assigned to these same options by any of the other theories in which we have some

[9] Ross, *Acceptance and Practical Reason*, 52-53, 57.

[10] Ross, *Acceptance and Practical Reason,* 400-403. (Note: See the final line of Theorem 1 while keeping in mind that it contains an abbreviation representing a complex term from line 14 and that this final solution is a summation requiring repetitive computation across all options under one theory - and then again and again for every theory being considered.)

credence.[11] In "Rejecting Ethical Deflationism," Ross refers to this move as assuming the option values assigned by different theories are "stochastically independent."[12] In his dissertation he formally expresses this stochastic independence by setting the statistical correlation between the option values delivered by any two theories to zero (0).[13]

The independence assumption allows Ross to drop typicality *entirely* from consideration since typicality just is the correlation of the values assigned by a theory to the values assigned by its complement. Setting correlation to zero is dismissing typicality from consideration. This move leaves Ross with only probability, selectiveness, and dispersion to consider. From there, he points out that, when selectiveness and dispersion are multiplied together (as would be the case when we seek their combined effects), the standard deviation appears in both the numerator and the denominator – because dispersion just *is* the standard deviation and selectiveness is the mean-maximum difference *divided by* the standard deviation. Multiplying the two allows the standard deviation in the numerator and the denominator to cancel out. Therefore, dispersion and selectiveness simplify into merely the mean-max difference. Ross labels this simplified max-mean difference as *"disparity,"* and he notes how his previously-complicated formula for determining the expected value of accepting a theory now has been reduced to the product of (probability x

[11] Ross, *Acceptance and Practical Reason*, 61.

[12] Ross, "Rejecting Ethical Deflationism", 767.

[13] Ross, *Acceptance and Practical Reason,* 403 proof line 1 of Theorem 2. (Note: There are additional confusing typos – this time in Ross's formal definitions D14 and D15 on pg. 401. In D14 he defines correlation between two theories T_A and T_B as $=E(z_{Ai} . z_{Ai})$ when he presumably means $=E(z_{Ai}.z_{Bi})$ if his goal is correlating two different theories. Also, in D15 he introduces his formulation of standardized maximum e_A (a.k.a. selectiveness) by erroneously referring to it as 'correlation' and using the correlation symbol. That this is error becomes clear when you view his linguistic definition and description of selectiveness on pages 50 and 51.)

disparity).[14] He expresses his concept most concisely in the final line of his appendix proof in "Rejecting Ethical Deflationism." In plain English it amounts to: the unconditional expected value of accepting theory A is greater than that of theory B if and only if the product of A's probability and disparity is greater than that of B's.[15] Voila! We have simplicity in the face of statistical sophistication.

I cheer Ross's quest to formalize a heuristic for theory selection, and I admire his creative insight in attempting to simplify the results. Yet, his simplifying move to dismiss typicality (by setting correlation to zero) can only be justified if the consequences of dismissing it are insignificant. They are not – for reasons I will provide shortly. But first, it's worth noting that Ross anticipates objections regarding the consequences of dismissing typicality from consideration. Recall that the effect of considering typicality is to disqualify theories that assign outlier values to options under consideration. This might be seen as disqualify fanatical theories; however, in what follows I will show how seeing typicality as *merely* disqualifying fanatical theories is a mistake.

Ross confronts – head on – the worry of fanatical theories. He acknowledges that, in the absence of typicality considerations, when we rely on high disparity, we enable consideration of fanatical theories. In other words, since high disparity is high difference between the theory's maximally rated option value and the mean of its option values, theories that assign a fanatically high value to an option become favored. Ross gives the example of a high disparity theory that posits wiggling ones toe as being supremely desirable over all other act options – thus devoting our efforts to toe-wiggling trumps

[14] Ross, *Acceptance and Practical Reason*, 62.
[15] Ross, "Rejecting Ethical Deflationism", 768.

striving for world peace as the most choice-worthy option. Yet, as he points out, Ross's simplified heuristic easily accounts for disqualification of such fanatical theories, since – for rational deliberators – the probability assigned to fanatical theories is exceedingly low. This, in turn, drives the product of probability and disparity to a level beneath that of more-plausible theories. That is, according to Ross, we do not need the typicality principle in order to avoid the consequences of fanatical theories under his heuristic – as long as we assume moral agents are rational beings who will assign probabilities reasonably. He is (pro tanto) correct. We are safe from overtly fanatical theories under his simplified heuristic when it is applied by rational agents, since those agents know enough to assign low credence to fanatical theories. Yet, avoiding fanatical theories is not the only reason to consider correlation (i.e. typicality) between competing theories. An additional reason is to avoid outrunning the tenable applicable range of *non*-fanatical, credence-worthy theories. In what follows I will explain the meaning of 'credence-worthy theories outrunning their respective ranges of tenability', and I will show how Ross's dismissal of typicality enfeebles his heuristic regarding detection of such overruns.

Consider utilitarian consequentialism along the traditional lines of Bentham, Mill and Sidgwick, whereby acts are morally right if and only if they maximize the good.[16] I will assume that the reader does not view utilitarian consequentialism as a fanatical theory since most of us have a considerable level of credence in the notion that we ought to consider maximizing good consequences as part of our action deliberations – even if this is

[16] Walter Sinnott-Armstrong, "Consequentialism", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/consequentialism/>.

not our main consideration. That is, consequentialism itself is not fanatical in the same sense as cherishing toe-wiggling is fanatical. Some libertarians might argue that *extreme proponents* of orthodox utilitarian consequentialism are fanatics. This may or may not be a true claim. However, even if it were true, we would not want to brand the theory itself as fanatical, and thus completely dismiss it from consideration during our moral deliberations. Rational, well-informed, well-meaning people may well have high credence in it, and – under Ross's simplified heuristic – they would assign it a high probability.

Additionally, consequentialism is a high disparity theory in regard to many decision scenarios. That is, in regard to many ethical deliberations, there is a wide difference between the maximal values it assigns to some options and the mean of the values it assigns. For example, it ranks giving my paycheck to a charity as being far higher than the average of the rankings of spending it on amusements, gadgets and other non-essentials. Given this high disparity, if I possess high credence in consequentialism, Ross's heuristic recommends accepting its prescriptions for giving much of my paycheck to charity.

Now, consider a more controversial prescription of consequentialism, namely: killing one human to save four. Peter Singer and Katarzyna de Lazari-Radek recently described the hypothetical scenario of a brain surgeon operating on a patient.[17] The patient is an ideal organ donor. In the scenario, there are four other dying patients who could all be saved by the organs of the surgery patient. Singer and de Lazari-Radek argue that, according to consequentialism, the brain surgeon ought to kill the patient being operated on and use his organs to save the four in waiting – *provided that the surgeon can do so*

[17] Katarzyna de Lazari-Radek and Peter Singer, "Secrecy in Consequentialism: A Defense of Esoteric Morality" *Ratio (new series)* XXIII 1 (2010) 34-58.

without being discovered and does not publically condone the act or publically promote acts of this nature.[18] This 'esoteric morality' is a coherent extension of utilitarian consequentialism provided that absolute secrecy of the act can be maintained. Singer and de Lazari-Radek note that secrecy is needed (in part) to avoid negative consequences regarding loss of faith in the medical profession if doctors are known to kill some patients to save others. They acknowledge that, in our real world, keeping these sorts of secrets is unlikely, and also that doctors who carry out such practices are likely to lead themselves into more risky moral judgment calls. Therefore, they admit we should advise doctors against such practices even though *the killings are the right thing to do*.[19] Singer and de Lazari-Radek go on to assert that esoteric morality in general should be disavowed in public while – in the same *published* paragraph – they paradoxically support it.[20]

Most of us want to resist the Singer and de Lazari-Radek conclusion. Suppose that a doctor publically admitted to having killed several patients in order to save many more. Would any of us claim that the wrong was in her publicizing her actions? Probably not! Among the expressions of outrage we would hear: "…sinful murderer!" or "…violation of a right to life!" or "…setting a horrific precedent!" or "…what kind of person could do such a thing?" In other words, the surgeon would have violated the taboos of the Mosaic code, libertarian inviolable rights,[21] the categorical imperative,[22] and virtue ethics.[23] Even

[18] de Lazari-Radek and Singer, 40.
[19] de Lazari-Radek and Singer, 41.
[20] de Lazari-Radek and Singer, 57.
[21] Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974) 30-33.
[22] "I ought never to act in such a way that I couldn't also will that the maxim on which I act should be a universal law." Immanuel Kant, *Groundwork for the Metaphysic of Morals,* (1785) Ch. 1 pg. 11. online with preface by Jonathan Bennett (2010) http://www.earlymoderntexts.com/pdf/kantgw.pdf (Note: We

though consequentialism ranks such secret killing as higher than fulfilling the Hippocratic Oath, the killings fit near the very bottom of at least four other widely-accepted, non-fanatical, theories. Yet, consequentialism is *not* – at core – a fanatical theory. It does not usually deliver such repugnant prescriptions, and many reasonable people have high credence in it – though most won't carry it to this extreme.

Let's consider whether Ross's heuristic will suggest rejecting the secret murder of one to save four. Suppose a non-philosopher divides his credence equally between aspiring to maximize good consequences and aspiring to avoid infringing on the rights of others. That is, he is split between honoring utility and honoring liberty, and he occasionally feels that annoying dissonance when the two collide. If we tell him about Ross's simplified heuristic, how might he evaluate the brain surgeon's choices? For simplicity, the surgeon's options are roughly: O1) refuse to operate – telling the patient the brain condition is inoperable, O2) operate and cure the patient, O3) operate and kill the patient but use the organs to save four others. The consequences are: C1) four die and one is left with brain disease, C2) four die and one is cured, and C3) one dies. If we determine disparity in units of human life, and we suppose the refusal to operate is worth approximately minus 1/2 life, then disparity is as follows: *(forced page break)*

might wonder whether the surgeon (or Singer) could will the killing as a universal law if their own daughter or son needed brain surgery.)

[23] "… we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts." Aristotle, *Nicomachean, Ethics Book II, paragraph 2* (350 BCE), translation W. D. Ross, available online at http://classics.mit.edu/Aristotle/nicomachaen.2.ii.html (Note: We can infer 'murderous by doing murderous acts.')

Table 1: Disparity of Utility vs. Liberty re Killing One to Save Four

Numerical units of human life	O1 Don't Operate & Don't Save the Four	O2 Cure the One & Don't Save the Four	O3 Kill the One & Save the Four	Mean	Max	Disparity (Max-Mean)
Utility	(-0.5)+(-4.0) = -4.5	(+0.5)+(-4.0) = -3.5	(-1.0)+(+4.0) = +3.0	- 1.7	+ 3.0	**4.7**
Liberty	- 0.5	+ 0.5	- 1.0	- 0.3	+ 0.5	**0.8**

Notice how, under liberty, the four transplant patients do not figure in the value equation for any of the three option scenarios, but they do under utility. The libertarian view is concerned with refraining from violating the equal rights of another, and the four transplant patients have no right to expect the doctor to kill another to save them – based on their own right to expect that the doctor refrain from killing them to save another. Therefore the deaths of the four patients by their own ailments are not a moral negative (regarding the doctor's decision) under the libertarian deliberation. Similarly, saving them is not to be awarded positive value if doing so involves violating the rights of another. However, under libertarianism we have an obligation to keep our promises since others have a right to expect delivery of what we freely promised (usually in trade). Therefore, under liberty there is negative value in the surgeon refusing to operate and positive value in her operating and delivering on her promise to do what doctors purport to do - i.e. heal.

The upshot of this exercise is to show that – if done in absolute secrecy so as to avoid the negative consequences of discovery – killing one patient to save four others has a *much*

higher disparity under utilitarian valuations than it has under libertarian valuations. Therefore, according to Ross's heuristic, if one has equal credence in utility and liberty, they ought to reject the libertarian view and support secretly killing an innocent to save four. Of course, using different ethical theories against utility will give different results. Some of them might end up with disparities higher than utility. However, it is not unusual for a person to hold utility and liberty as their two, highest-credence theories. My point is that, in some instances, Ross's heuristic will fail to recommend rejecting an ethical conclusion that most of us abhor.

First objection! A Ross defender may ask, "So what? Where's the foul?" More specifically, they might claim that Ross's heuristic is doing what it sets out to do. Someone with equal credence in utility and liberty needs a reasonable means to adjudicate between the two theories, and Ross's heuristic provides such a means based on a desirable attribute of theories in general – i.e. disparity. We might find the prescriptions of the favorably-adjudicated theory unsettling, but that finding is not evidence that the heuristic has failed to reveal which ethical theory the agent has most reason to accept.

I reply that our abhorrence *is* evidence of heuristic failure, because it is symptomatic of having failed to adequately address a key concept that Ross initially attempted to capture with his typicality principle, namely outlier capture. Our finding the secret killings so abhorrent is evidence of the heuristic failing the outlier capture concept. I assume you believe that you do not deserve to die under a surgeon's knife merely because someone else became ill through no fault of your own. If an ethical theory prescribes such a death for you, the blatant unfairness of the prescription is evidence of an outlier. Recall that

Ross did *not* back away from the outlier capture concept when he dismissed typicality. He, instead, suggested that the probability principle will do some of the work in capturing outliers by rejecting fanatical theories. Yet here is an instance of an outlier prescription having slipped through the heuristic. It slipped through because the theory that generated it is not – at core – fanatical enough to be dismissed under the improbability principle.

Of course, the agent in our hypothetical example has credence in only two theories, whereas you or I might have credence in several more. That is, the outlier status will be more obvious to us. I have not yet established whether Ross's *original* formal version of the typicality principle will capture the outlier prescription under this particular scenario. (I will address that question shortly.) Nonetheless even our two-theory agent is likely to see the prescription for secret killings as evidence of heuristic failure regarding outlier capture – particularly if it is his turn for the surgeon's knife. At this stage of my argument, the point to take away is that the probability principle – as formulated by Ross – is not enough to capture some abhorrent outlier prescriptions.

Given the above, we have reason to believe that Ross has traded away an important theory selection concept (i.e. outlier capture) when he accepted stochastic independence of value assignments between theories, and thereby erased typicality from consideration. He did so in the name of simplification, and offered justification by describing how the probability principle prevents rational agents form assigning high credence to fanatical theories which, otherwise, would require typicality for their rejection. Yet, Ross did not consider whether the probability principle shields us from outlier prescriptions delivered by non-fanatical theories. It does not. Recall that the probability principle holds we have

pro tanto reason to accept theories we regard as more probably true than theories we regard as less probably true. Thus, action guiding based on subjective probability is limited to theories as wholes and not their prescriptions – according to Ross.

Based on what has preceded, there is not only be good reason to hold on to typicality, but also good reason to widen the scope of probability to include individual prescriptions. By accepting a prescription for secret killings, Ross's heuristic tacitly endorses a view whereby ethics philosophers who favor liberty, or virtue, or the categorical imperative are simply mistaken. It also implies that folk moral intuitions abhorring secret murders are naively incorrect. I am unwilling to accept a heuristic that endorses such a view because I find it highly improbable that so many of our ethical theories and moral intuitions could mistake a *most-choice-worthy good* for a *widely-held taboo*.

Second objection! A Ross defender might claim I am misrepresenting Ross, because the heuristic output is not a truth claim regarding the accepted theory. They might even point to the fact that I acknowledged as much when I introduced Ross's probability principle at the outset of this paper. They might emphasize that the heuristic is meant to be a guide when we have credence in more than one theory and we don't know which is true. That is, the heuristic is only a practical reasoning tool aimed at helping us find what we have most reason to choose. Therefore, it is inappropriate for me to criticize the heuristic for supposedly mistaking a moral taboo for a choice-worthy good since the heuristic makes no claim as to the truth status of the accepted theory's recommendations.

In reply, I assert that the above characterization of Ross's project is not quite accurate. While Ross is careful to state that his heuristic outputs are not outright truth claims about

the accepted theories, this does not mean his project is divorced from truth quest. In fact, *the very foundation of the heuristic rests on assumption of truth regarding the accepted theories*. Consider how Ross founds his project when he defines 'accept a theory'.

> "By *to accept a theory*, in relation to a given decision problem, I mean *to guide one's decision on the basis of this theory.* More precisely, to accept a theory is to aim to choose whatever option this theory would recommend, or in other words, to aim [to] choose the option that one would regard as best on the assumption that this theory is true." – [emphasis in original, grammatical correction added.][24]

From this passage it is clear that Ross intends his theory-acceptance heuristic to involve an assumption that the accepted theory is true. Thus, it is fair for me to offer criticism pointing to – what most people see as – a probable mistake regarding the truthfulness of the prescriptions made by a theory accepted via the heuristic. Also, the words "to choose whatever option this theory would recommend" make clear that Ross intends acceptance of a theory under his heuristic to mean acceptance even of seemingly improbable prescriptions made by that theory. This places the burden of weeding out improbable results on the ability of the heuristic to reject theories that deliver such results. Again, this is precisely the reason why a theory-acceptance heuristic must include a sufficient means to reject outliers, and why Ross should not have eliminated the typicality principle when he moved to simplify. If a sufficient means to reject outliers is not to be provided within the algorithm of the heuristic, then the domain of the probability principle must be modified to extend to individual prescriptions of the theory.

[24] Ross, *Acceptance and Practical Reason,* 5.

Third objection! An antagonist may complain that, by suggesting we allow for rejecting theory prescriptions on the basis of subjective improbability, I am allowing agents to pick and choose outputs based on mere seeming of prescriptions. In other words, I am defeating the reason to have and use ethical theories in the first place, namely: to provide some form of axiomatic or rule-based guidance for actions that have moral consequence. Thus, the application of the probability principle belongs at the level of assessing the theory's likelihood, not the likelihood of its prescriptions.

I would be inclined to agree – if only I could muster confidence there exists an actionable ethical theory that always delivers truth. I have no such confidence because – in regard to ethical theory development – we've been at it for thousands of years, and our best thinkers are still widely divided in regard to claims about the one true theory. This observation points to the very reason for considering Ross's project in the first place. I will address that reason in the concluding remarks to follow shortly. But first, I will offer a specific redirection of the objection. I assume we agree that agents should reject highly improbable theories (e.g. toe wiggling) in favor of more probable theories (e.g. utility, liberty, virtue, etc.) Improbable theories deserve rejection because the truth-consistency of their prescriptions is improbable. Our goal is true prescriptions – or, at least, probably-true prescriptions. Therefore, it is both goal-defeating and unreasonable to treat improbability as a reason to reject a theory, but not as a reason to reject a prescription – unless, of course, we *know* a particular theory is always true, in which case we have no need of a heuristic.

To be clear, in what has preceded, I do not mean to single out utilitarianism as untenable. It is, after all, one of our most widely accepted ethical theories, and its

prescriptions are not usually as abhorrent as secret killings by surgeons. The libertarian view suffers similar difficulty when it is pushed to its limits. We can readily devise hypothetical, libertarian-driven examples that prescribe results simultaneously abhorrent when viewed from the perspective of several other theories. The goal is not to brand any of our reasonable (and helpful) ethical theories as being 'fanatical'. It is to identify when those theories are not good choices for the circumstance at hand. *Every* ethical theory (so far) delivers abhorrent prescriptions under at least some circumstances – as is evidenced by the persistent moralistic debates about correct theory among ethical philosophers, theologians, and politicians. We are fortunate in that, when circumstance causes a given high-credence, non-fanatical theory to deliver a seemingly untenable prescription, there are several other credence-worthy, non-fanatical theories that we can compare it to. These comparisons help determine whether the particular repugnant prescription is merely an outlier. The ability to make them is crucial to the justification of our assumption that the accepted theory is true – in regard to the options under consideration.

Fourth objection! A Rossian might complain that my hypothetical example was about an individual who had equal credence in liberty and utility, yet I am pointing to the prescriptions of many additional theories to disparage the results of Ross's heuristic regarding this two-theory example. Thus, the objector might continue, even if Ross had not made his simplifying assumption by eliminating typicality from consideration, the statistical goal of discovering outlier prescriptions looks to be highly unlikely with only two theories under consideration. I completely agree. However, this observation in no way refutes my assertion that the typicality principle should not be removed – since many

of us *do* hold credence in three or more theories. Nor is this objection a defense of Ross's heuristic – simplified or not. It is, instead, evidence of another shortcoming, namely: *ineffectiveness under bi-elemental theory sets*. That is, even if typicality is taken into consideration under the original, non-simplified Rossian heuristic, that heuristic is ineffective at discovering outlier (i.e. fanatical) prescriptions when the agent's set of credence-worthy theories has only two elements, because the complement of each credence-worthy theory is a single-element set. Here is why that's a problem.

Consider again how typicality works. The typicality of a given theory under a given set of options is found by first calculating the statistical correlation between the option values assigned by that theory and the weighted average of the option values assigned by that theory's complement. Then the correlation of each theory is compared to the correlation of each other theory. Theories with higher correlation are preferred to those with lower correlation, since the lower the correlation to the group the more the theory is an outlier – regarding the particular option set under consideration.

The complement to a theory is the set of all *other* theories under consideration - i.e. all of the other theories in which the agent has credence. Think of the complement itself as an alternate theory comprised of many theories weighted by respective credence. So, the correlation between a theory and its complement is the correlation between two theories – one of which is a compilation of many theories. Ross defines correlation between two theories (A, B) using definition D14 in appendix C of his dissertation:[25]

[25] Ross, *Acceptance and Practical Reason,* see D14 and D15 on 401. Also, for clarity, note the D14 and D15 error corrections mentioned in my footnote 13 above.

D14. $\rho_{A,B} = E(z_{Ai} \times z_{Bi})$

In English, D14 means the correlation of theory A to theory B is the expectation for the expression of: [(z for the value assigned to option i by theory A) times (z for the value assigned to option i by theory B)]. The 'E' expectation simply means that the calculation in the brackets must be done for each option i under consideration, and then the average of all of those calculations must be found to give the expectation. The nature of z in the expression is unimportant in regard to the point I wish to make. (For completeness, it is the value of a given option assigned by the theory minus the average of the values assigned by that theory, and that difference is then divided by the standard deviation of the values assigned by the theory.)[26] The important thing to note about z is that z_{Ai} and z_{Bi} are multiplied together in the D14 expression, and multiplication is a symmetric relation – i.e. $(a \times b) = (b \times a)$. This symmetry is a key to what immediately follows.

The correlation between a theory and its complement – by itself – is of little value until we compare it to the correlations of each of the other theories under consideration. Each of those other theories has its own complement formed by every theory in the group *except* that theory itself. Thus, when we have *only two* theories in the group, each is the entire complement of the other. Under this circumstance, the computed correlation of the first theory will be exactly the same as the computed correlation of the second. Here's why:

Consider the group of theories composed only of Utilitarianism (U) and Libertarianism (L):

[26] Ross, *Acceptance and Practical Reason,* see D9 on 401.

(1) Let G = the theory group comprised by U and L such that G = {U, L}

(2) Let Uc = the complement of U; therefore, Uc = {G – U} = {L}, i.e. Uc = L

(3) Let Lc = the complement of L; therefore, Lc = {G – L} = {U}, i.e. Lc = U

(4) From D14., $\rho_{U,Uc} = E(z_{Ui} \times z_{Uci})$

(5) Also from D14., $\rho_{L,Lc} = E(z_{Li} \times z_{Lci})$

(6) However, from (2), Uc = L, so L = Uc

(7) Therefore from (5) and (6), $\rho_{L,Lc} = E(z_{Uci} \times z_{Lci})$

(8) Also from (3), Lc = U; therefore (7) becomes $\rho_{L,Lc} = E(z_{Uci} \times z_{Ui})$

(9) However, under multiplication symmetry, (8) becomes $\rho_{L,Lc} = E(z_{Ui} \times z_{Uci})$

(10) Therefore from (4) and (9), $\rho_{L,Lc} = \rho_{U,Uc}$

In other words, when liberty and utility are the only two theories in which one has credence, the correlation of liberty is exactly equal to the correlation of utility. This means that the typicality principle has no force under this circumstance. Recall that typicality is determined by comparing correlations. Theories with higher correlations to the group are more choice-worthy in regard to the options at hand than theories with lower correlations – all things being equal. Since the two correlations in our example are equal, neither theory can be determined as more typical than the other in relation to the bi-elemental theory set in which the agent has credence. This result applies generally to *any* theory group containing only two theories. However, this result would not obtain if a third theory were added (say Virtue = V); since G would then equal {U, L, V} and the complement Uc would be {G – U} = {L, V} instead of only {L}. This foils the substitutions that led to the line (10) equality conclusion.

The upshot of these considerations is that typicality requires consideration of three or (preferably) more theories before it can do any work adjudicating between theories. Yet typicality is the key means to discover when normally acceptable, non-fanatical theories have been exposed to an option set that causes them to deliver abhorrent (i.e. fanatical) results. Therefore, even when we add the typicality principle back – thereby returning Ross's theory to its original state – there is still a serious shortcoming. Person's who only have credence in two theories will not be able to detect when one of those theories is giving outlier prescriptions – at least not via Ross's heuristic.

Allow me to summarize the results of this investigation thus far. In what has preceded, I have shown how the simplified version of Ross's heuristic is unable to detect some abhorrent outlier prescriptions delivered by normally choice-worthy, non-fanatical theories because his reliance on the probability principle is limited to considering rejection of improbable theories but not improbable prescriptions. I have asserted that a viable theory acceptance heuristic should extend the domain of probability consideration to prescriptions of theories rather than being limited solely to the theory axioms. I have also shown how Ross's simplification assuming stochastic independence enfeebles his heuristic in regard to its ability to capture and reject theories giving outlier prescriptions when an agent has credence in three or more theories. Additionally, I have uncovered a serious limitation of the non-simplified version of Ross's heuristic in regard to agents who have credence in only two theories – i.e. bi-elemental theory sets.

How might the exposed issues be resolved? One obvious path is to give up the assumption of stochastic independence and return the typicality principle. This still leaves

open the bi-elemental theory set problem. The bi-elemental issue might be solved by stipulating that two-theory agents must temporarily 'adopt' additional theories for deliberation purposes. This solution is somewhat analogous to detectives adding three or more known innocents to a criminal line up when they have only two credence-worthy suspects. However, this solution requires an additional complication in that we would need to further stipulate that the 'adopted' theories must come from a pool of theories somewhat similar to the agent's subjectively credence-worthy theories – e.g. no toe wiggling. Of course, even if we could make this adoption suggestion work, the resulting heuristic is far from useful for practical action guidance because of its unwieldy, calculation-intensive nature.

By way of example, I have performed the correlation calculations to determine typicality using the options values presented in Table 1. I added virtue theory into the mix to avoid the be-elemental set issue, and I assumed the virtue theory value ascriptions would match those of liberty. After more than an hour of tedious work and considerable wear on my pencil eraser, I managed to generate one and one half pages of calculations, and I found that – as was expected – the utility theory values were less typical than those of liberty or virtue values. However, the values differed by only 5.5% of the full range of correlation (-1 to +1), due to the fact that I used only three theories. These exercise results reinforce the need to have *many* (i.e. five or so) theories under consideration in order for typicality to reliably resolve outliers. Considering five or more theories will require four or so pages of calculations, thereby exacerbating the unwieldy nature of the heuristic even further.

I conclude that Ross's heuristic is impractical for practical action guidance when employed in full form – i.e. with typicality and with the required stipulatory adoptions. Alternately, its findings are untrustworthy when it is employed with the enfeebling simplification of stochastic independence. I am uncertain whether there is a way to simplify the full-form heuristic; however, there is good reason to try and good reason to subject Ross's heuristic – or something like it – to further consideration. Given the tremendous historical investment in moral theory and the still-uncertain status of its products, we have reason to consider whether moral truth is analogous to logical truth. Kurt Gödel has taught us that no single consistent axiomatic logic system is complete.[27] That is, no consistent axiomatic system can uncover all logical truth. Of course, we don't know if moral truth is like logical truth. Moral truths might simply be a individual – sometimes conflicting – contingent features of the world, or perhaps elements in a rule-set specified by deity, or perhaps no truth at all but practical social conventions. Despite this uncertainty in the nature of moral truth, analytic philosophers treat it as if it were analogous to logical truth by persistently founding their theories on axioms and building theorematic prescriptions from those axioms. If we expect to illuminate moral truth with this type of treatment, we ought to also expect that each of our consistent axiomatic theories will display incompleteness – i.e. will miss some truths. Therefore, those thinkers

[27] Roger Penrose, *Shadows of the Mind – A Search for the Missing Science of Consciousness* (Oxford: Oxford University Press, 1994), 90.

who are committed to taking an axiomatic approach to discovering moral truth have good reason to view a Rossian-type heuristic as a worthy pursuit.[28]

Bibliography

Aristotle. *Nicomachean, Ethics Book II, paragraph 2* (350 BCE), translation W. D. Ross, available online at http://classics.mit.edu/Aristotle/nicomachaen.2.ii.html

de Lazari-Radek, Katarzyna and Peter Singer. "Secrecy in Consequentialism: A Defense of Esoteric Morality" *Ratio (new series)* XXIII 1 (2010) 34-58.

Kant, Immanuel. *Groundwork for the Metaphysic of Morals* (1785) online with preface by Jonathan Bennett (2010) http://www.earlymoderntexts.com/pdf/kantgw.pdf

Nozick, Robert. *Anarchy, State, and Utopia* (New York: Basic Books, 1974).

Penrose, Roger. *Shadows of the Mind – A Search for the Missing Science of Consciousness* (Oxford: Oxford University Press, 1994).

Ross, Jacob. "Rejecting Ethical Deflationism", *Ethics* 116 (July 2006): 742-768.

Ross, Jacob. *Acceptance and Practical Reason,* (Ann Arbor: ProQuest Information and Learning Company, 2007).

Sinnott-Armstrong, Walter. "Consequentialism", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/consequentialism/>.

[28] Ross, *Acceptance and Practical Reason*, on page 245 Ross considers how Gödel's incompleteness can sometimes block subjective ought regarding rational choice.