prometheus

The Johns Hopkins University Undergraduate Philosophy Journal

2015

prometheus

2015

prometheus
The Johns Hopkins University Undergraduate Philosophy Journal

Prometheus is an international undergraduate philosophy journal published by students at the Johns Hopkins University. The purpose of the journal is to promote philosophic discourse of the highest standard by offering students an oppurtunity to engage in open discussion, participate in the production and publication of an academic philosophy journal, and to establish a community of aspiring philosophers from all walks of life.

www.prometheus-journal.com

Contents

To Kevin Powell for the countless hours he devoted
to this journal and its citations.

To David Lindeman for his guidance and mentorship over the years.

To David Williams, Virgil Demario, and Dan Friedman
for their edits and support.

To Carlo Olcese, Tracy Kao, and Edwina Picon whose hard work and
artistic vision have brought these pages to life.

To our professors who inspire and support our love of philosophy.

And, finally, to our readers. Philosophy is meant to be shared.
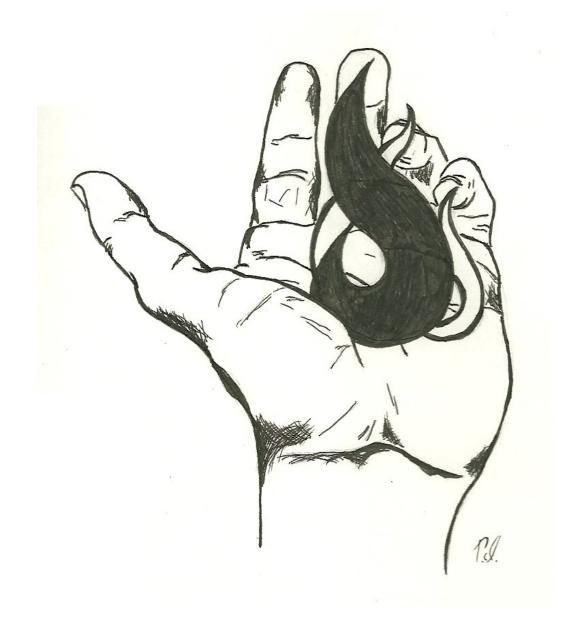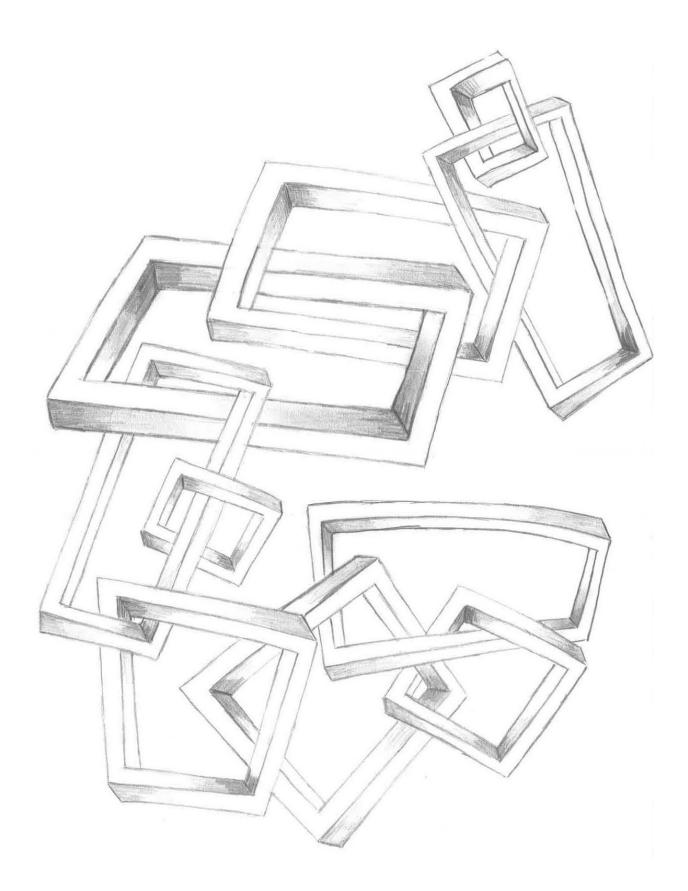
We dedicate this to you.

As always, Prometheus is proud to present this year's publication, which comprises the highest-quality submissions we received in 2015. Papers were written by students across the globe, a select few of which made it through review by our undergraduate staff readers, executive board, and graduate reviewers. Here, we present three papers whose topics not only range temporally, from Plato to philosophers of today, but also cover multiple sub-disciplines, from metaphysics to ethics.

The success of nearly every academic field, from the natural sciences to history to sociology, depends on the testing, tuning, and consequent improvement of theories. Philosophy is no exception. Though its methods may differ notably from those of, for example, the lab sciences, the principle behind these methods is the same. For the most part, instead of empirical tests, philosophers employ logic to critique and develop one another's theories. After all, according to philosopher Massimo Pigliucci, "You can think of philosophy as an exploration of conceptual, as opposed to empirical, space, concerning all sorts of questions ranging from ethics to politics, from epistemology to the nature of science."

Without journals, philosophy would lose a crucial tool for the advancement of theories: the exchange of ideas between the author and her critics. Moreover, we believe that it is important to start the philosophical conversation at an early stage. Great discussions and theories oftentimes originate from our fellow undergraduates and peers. That is why we are proud to offer them a platform to share these ideas, which not only serves to further important debates concerning ethics, language, science, and other topics, but also allows our authors to gain experience in the philosopher's "laboratory."

It has been an honor and privilege to helm such a fine journal. We are resolved to make it still better in our years to come.

Best,
Cameron Davis and Helen Zhao
Editors-in-Chief, *Prometheus*

2015

COMBATANTS IN A CORNFIELD
Matthew B. Gibbons
Rutgers University, Class of 2016

ABSTRACT

This essay initially provides a thought experiment regarding moral permissibility. The situation depicted is named Combatants in a Cornfield. Combatants in a Cornfield realistically represents war time scenarios which have taken place since the beginning of America's War on Terror. The essay then explores a few of the dominant views of permissibility and shows how their applications to the realistic scenario given are limited. The limitations stem from being fact-relative, and the result of their fact-relativity is that they cannot be used by the subjects in the scenario to guide actions. Finally, the essay identifies key aspects necessary for a moral permissibility theory which can be used by subjects with limited evidence to guide action, and formulates and argues for a theory which does satisfy those requirements.

TEXT

The following scenario is both realistic and typical of modern combat. Consider the case of:

"Combatants in a Cornfield"

A group of combatants, specifically a squad of American Marines, is patrollingc the area around their base in Afghanistan. That area is defined by farmland sprawling for miles in every direction randomly interspersed with small villages. Some farmers are tending to their crops. Occasionally small children approach the patrol to ask the Marines for candy. Nonetheless, the area is known to have enemy combatant groups in it. In order to avoid a minefield, the Marines walk through a cornfield. Once they reach the center of the cornfield, the Marines begin to take machine gun fire. They are immersed in tall cornstalks restricting their knowledge of the situation. What they know is that there is a direct threat. They can approximate the position of each of their squad mates. They can hear distant enemy gunfire coming from a general direction. The precise location, strength, and nature of the threat are uncertain. They see bullets impacting, kicking up soil, and cutting through the corn around them and their friends.

It is possible that civilians—farmers and children—are in the field taking refuge from the gunfire. Should the Marines resign themselves to be killed in that cornfield, or should they return fire in the general direction of the threat in order to eliminate it and possibly kill the farmers and children who may be nearby? Do they have other options? By their reckoning, any attempt to flee would be clearly visible to the enemy, by movement of the cornstalks, and may be tantamount to resigning themselves to death. What should they do?

Neither the prevalent views of the morality of defensive killing offered by Jeff McMahan, Judith Thomson, and others, nor the traditional Just War Theory, offer clear guidance to the Marines. That is extremely surprising considering that scenarios such as the one above are common in modern warfare and that McMahan and others believe that ethics in the conduct of war, including scenarios such as this, is essentially reducible to the ethics of individual self-defense. Therefore, these theories purport to differentiate right actions from wrong actions in cases like Combatants in a Cornfield, but the guidance that they provide the actors in the scenario is unclear. In retrospect, for the purposes of evaluating actions and making moral judgments about them, the theories are applicable and provide answers as to what was right, wrong, permissible, or not. But when faced with ambiguity and without all the relevant facts, the theories are unable to provide answers looking forward instead of back. They are unable to provide actors in the present with guidance about the permissibility of their potential, or future, actions.

In this paper, I argue that there is a form of evidence-relative moral permissibility that has the unique ability to guide the actions of people faced with a moral predicament. Moral predicaments may vary in terms of their severity and frequency. It is likely that in war, a combatant faces increases in both their likelihood of facing a major moral predicament at all and the frequency with which that will occur. In this essay I will focus on the Marines facing the Combatants in a Cornfield scenario above. That scenario typifies the reality of combat in wars today. By comparing the factors present in the Combatants in a Cornfield scenario to the factors necessary to utilize international law and moral theories in order to judge moral permissibility, we will see how the law and each of those theories lack the ability to guide the Marines' actions. Then we will examine what the problem is that prevents them from guiding actions. I will define the relevant terms before exploring attempts to solve the problem. Finally, I will provide a combination of principles, which, I believe, does solve the problem. That combination of principles provides a theory of permissibility that is applicable to real life and can be used to guide action.

The law governing the Marines' actions and the theory it is based on do not practicably answer the question, "What should the Marines do?" International humanitarian law, known as the Law of Armed Conflict (LOAC), is similar to the principles of traditional Just War Theory (International Committee of the Red Cross, *Commentary* 33-56). Both LOAC and Just War Theory are able to tell the Marines what is morally permissible only if the Marines have all the relevant facts. The Marines do not have all the relevant facts, but nonetheless they must act. What they need is guidance for their actions, a framework that will tell them how to act and what is morally permissible or impermissible even in the absence of relevant facts. Neither LOAC nor the traditional Just War Theory is able to provide action guidance under the realistic condition of restricted factual knowledge.

Of the principles of Just War Theory, the ones that might apply to the Combatants in a Cornfield scenario are necessity and distinction. The principle of military necessity requires that in a just war, just combatants must attack only military objectives that will contribute to the defeat of the enemy. The principle of military necessity in the case of Combatants in a Cornfield is clearly satisfied by virtue of the self-preserving defensive position the Marines are in. The principle of distinction, however, is problematic for the Marines in the cornfield because they are unable to distinguish between

(a) the combatants whom they are authorized to defend themselves against and (b) the noncombatants who may also be in the area (Walzer 4-5). The principle of distinction requires just combatants to commit acts of violence only towards enemy combatants and never towards noncombatants stuck in circumstances that they did not create. In Combatants in a Cornfield, the possible presence of noncombatant civilians who are caught in the situation due to no fault of their own, combined with the Marines' reduced visibility, prevents the principle of distinction from being clearly satisfied. The traditional Just War Theory principles, then, do not help the Marines decide what it is morally permissible to do.

The laws that the Marines in Combatants in a Cornfield are subject to are equally unhelpful. LOAC requires that the Marines shoot only at the enemy combatants who are actually targeting them (Haque 71-71). Since the Marines cannot see past the cornstalks, they are unable to tell who is targeting them. Due to the imprecise nature of LOAC, it is unable to deal with the uncertainty the Marines face (Haque 66).

The standard views of killing in self-defense within the philosophical literature are also unable to answer the question. One such view, proposed by Jeff McMahan, is known as the "liability to defensive killing" account. McMahan's view holds that it is in fact permissible to kill the attacking

enemy because she is liable to be killed in self-defense. However, McMahan would claim that it is impermissible to kill any non-combatants who may be nearby, because they are not liable to defensive killing (*Basis of Moral Liability* 394-401).[1] Another account of killing in self-defense is Thomson's rights-based account. Like McMahan's view, Thomson's categorically allows for the Marines to kill the attacking enemy aggressors. Also like McMahan's, it definitively prohibits the killing of the noncombatants. Because the noncombatants, who may or may not be present, have not violated anyone's right-not-to-be-killed, they have not forfeited their own. Thus, Thomson would claim that killing them is impermissible (284, 290-295).

If the Marines of Combatants in a Cornfield were to apply either of those accounts to their situation, then killing the aggressor combatants is certainly permissible. Also, killing the noncombatants is certainly impermissible according to both accounts. The Marines in the scenario cannot distinguish between combatants and noncombatants; thus, they are unable to differentiate between those whom it is permissible to kill and those whom it is not. They are facing their own impending deaths and seek guidance for the permissibility of their actions, but neither McMahan's nor Thomson's permissibility theories can help the Marines decide what to do. Because of the limited factual knowledge available to the Marines,

the same uncertainty that characterizes modern warfare and many other kinds of real-world moral decisions, these two dominant permissibility theories fail to guide action.

Jonathan Quong's account, known as the Principle of Defensive Killing (PDK), may be the most helpful and applicable theory of permissibility. Quong's PDK holds that:

> You can permissibly infringe someone else's rights [such as killing them] when this is necessary to defend something that belongs to you [such as your own life] provided: (a) the infringement is proportionate, and (b) you would be able to keep what belongs to you if the other person and all their property were not present. (530)

The first requirement of PDK, (a), is satisfied because the Marines themselves are under threat of death, and thus responding by killing is proportionate. Additionally, the Marines have no reasonable method by which to save themselves other than by killing the aggressors. Here, an assumption is required: if the Marines fled, that fleeing would result in their deaths. That assumption is reasonable. A retreat would result in movement of the cornstalks, and the machine gun fire would likely be adjusted to continue to accurately target them during their movement. Further, the second requirement of PDK, (b), is

satisfied because if the machine gunner were to be killed, the threat would be eliminated and the Marines would no longer face a direct and immediate risk of losing their lives. Thus, it appears the Principle of Defensive Killing provides clear guidance: returning fire and killing the enemy machine gunner is in fact permissible. But Quong claims that under certain circumstances a self-defender forfeits her permission to kill in self-defense, and doing so becomes morally impermissible. One possible scenario in which one forfeits the right to kill in self-defense is when one kills bystanders (Quong 518-520). So the Marines, unable to identify the machine gun's exact location and to differentiate between possible civilians and combatants, are also unable to apply the PDK in deciding what they should do.

According to the International Committee of the Red Cross, the intent of LOAC is to restrict and guide the actions of participants in order to "limit the effects of armed conflict" (International Committee of the Red Cross par. 1). Similarly, the intent of theories of moral permissibility may be to create a framework for reference in making decisions with moral ramifications. Haque portrays the existing international laws as contingent upon specific knowledge and unusable without it (66-67). The theories of permissibility I have discussed are fundamentally unsuited to situations they are intended to apply to. LOAC and each

view of permissibility listed above attempt to show what the Marines can permissibly do. In practice, though, each is dependent upon some aspect of knowledge that the Marines cannot obtain. Since the Marines do not have access to the prerequisite knowledge, neither LOAC nor any of the listed moral views apply to them in a practicable way.

Therein is the problem: regardless of intent or design, the Marines are unable to use any of these theories to guide their actions. The root of that problem is that all of these theories concern what Derek Parfit calls the "fact-relative sense" of moral permissibility (143). In other words, they tell us what it would be impermissible to do if the Marines knew all the relevant facts. Each theory of moral permissibility above is fact-relative.

Parfit distinguishes fact-relative permissibility from evidence-relative permissibility. An act is wrong in the evidence-relative sense when the act would be wrong if the relevant facts were what the available evidence gives us sufficient reason to believe they are. Parfit's final sense of permissibility is the belief-relative sense, in which an act is wrong if it would be wrong if our beliefs about the facts were true (143).

Relating to both evidence-relative and belief-relative permissibility is the category of permissibility known as subjective rightness, which is recognized in the literature by

philosophers such as Holly Smith and McMahan. Smith might summarize her views and define subjective rightness as the property of an act that makes it seem to be right to the subject based on their beliefs about the facts of the case at the time of the act (Smith 72-82, 92). McMahan similarly might characterize subjective rightness as the objective permissibility of an act if the subject's reasonable beliefs about the facts regarding that act were all true (Jeff McMahan, *Torture* 103, 106-107).

Subjective rightness and evidence-relative permissibility are examined far less often than fact-relative permissibility throughout the philosophical literature.[2] McMahan concedes, though, that for at least some people who have potentially acted in an objectively wrong manner that was subjectively right, "there was no rational basis available... for doing otherwise than [they] did" (*Torture* 107). It appears, then, that for all the study and analysis of fact-relative permissibility, subjective rightness is related most directly to the "rational" decision-making people perform every day (McMahan, *Torture* 107).

The accounts of standard Just War Theory, LOAC, McMahan's liability account, Thomson's rights-based account, and Quong's PDK are all fact-relative theories, yet they are supposed to determine permissible and impermissible conduct in cases such as my Combatants in a Cornfield. When any fact-relative permissibility accounts

are applied to situations with inherent uncertainty, such as Combatants in a Cornfield or any real scenario, they fail to be useful in a specific way. These fact-relative accounts fail to tell agents everything they need to know in order to make moral decisions. Guiding actions and contributing to the decision-making process is the ultimate goal, and that is where the utility of permissibility theories lies. The accounts listed require facts that are not evident, apparent, or even accessible to the situation's participants. For that reason, they fail to guide action.

The ability to guide action cannot depend on a total knowledge of the facts of a situation. While thought experiments can stipulate knowledge, in real situations, knowledge is often restricted to what is apparent. Choices in thought-experiments can be defined by their consequences, such as "either one life is lost in this case, or three lives are lost in that case." Choices in reality are made with restricted knowledge and can be between options with unspecified results. Fact-relative theories of permissibility have no direct use in those cases because the facts that are required are not known.

In Combatants in a Cornfield, the Marines looking for guidance need an evidence-relative theory of moral permissibility to aid their decision-making process. In order to be useful in guiding decisions, whether in war or not, an account of moral permissibility must be evidence-relative. Evidence-

relative permissibility is uniquely equipped for application to real-life decision-making because real-life decision-making involves uncertainty. Evidence-relative permissibility can be applied even when subjects have only limited access to the facts of a situation. The range of access to a situation's facts, from very little access to almost total access, results in varying levels of certainty about the effects of an act in that situation. The chances of those effects occurring due to the act, multiplied by the goodness of the effects of that act, is what Parfit calls the "expectable goodness" of that act. For any situation, whichever act has the most expectable goodness can be called "expectably-best" (Parfit 150).

Parfit provides an account of evidence-relative permissibility that he calls "expectabilism". Expectabilism is the belief that "[w]hen the rightness of some act depends on the goodness of this act's effects or possible effects, we ought to act, or try to act, in the way whose outcome would be expectably-best" (Parfit 150). The expectabilist view can be applied to the Marines in the Combatants in a Cornfield situation with multiple, and contradictory, outcomes. On one view, it would be expectably-best for the Marines to ensure their own survival by returning fire in an attempt to kill the enemy force targeting them and potentially killing the civilians near that enemy force (Parfit 150). That view requires a value-judgment: the lives of the Marines are worth more than the lives

of both the enemy and the probable number of civilians killed. The Marines have finally obtained guidance for their actions. With the previous required judgment and an expectabilist view, it is morally permissible for the Marines to return fire and attempt to kill the enemy. However, if that value-judgment premise is rejected, then the permissibility of that act changes. If the probable number of civilians killed is likely to outweigh the value of the Marines' lives, then that expectabilist account states that it is morally impermissible for the Marines to return fire, and the guidance for their actions is reversed. In that case, expectabilism requires them to act in the opposite manner. The cause of the reversal is a value-judgment for which the Marines have no guidance because they have no means of assessing it—the number of lives at stake is just unknowable to them.

The expectabilist account gives the Marines two contradictory action-guiding answers to the question of what to do, creating what Smith calls a "paradoxical tension ... by evaluating an agent's action as both right and wrong" (72). The difference between the two moral judgments is once again knowledge that is unobtainable for the Marines. Rather than address the problem of an unobtainable requirement, Parfit analyzes the various decisive reasons for acting by discussing different definitions of "ought" (152).[3] At least in cases of war, such as the Combatants in a Cornfield scenario,

there must be a better solution that can provide the Marines with an action-guiding theory of moral permissibility.

The ability of expectabilism to guide action is dependent on the answer to a question: what is the expectable probability that the lives of the civilians are worth more than the lives of the Marines? That question is actually a conflation of two questions:

> *Question (a): What is the number of civilian and enemy combatant lives at stake?*

Question (a) has a definite answer, but it is not known and it is unable to be known by the Marines. And

> *Question (b): What is the value of those lives compared to the lives of the Marines?*

It is markedly apparent that in the decision-making process in the real world, evidence-relative permissibility is the kind of permissibility that matters. That is simply because people are not omniscient and must act with imperfect knowledge. Fact-relative theories of permissibility, which are dependent on complete factual knowledge of a situation, are therefore unable to guide action. Action-guiding frameworks such as international laws governing use of violence in war and many leading theories of moral permissibility are fact-relative. Those frameworks are ineffectual because they require complete knowledge of

facts that no subject can obtain. If there is going to be an effective theory of permissibility that can be used to guide action, it must meet three criteria. It must (1) be evidence-relative, so that people in real situations with knowledge constraints can utilize it; it must (2) answer or provide a method of avoiding the problems posed by Question (b); and it must (3) answer or provide a method of avoiding the problems posed by Question (a).

Expectabilism meets criterion (1) but gives rise to criteria (2) and (3) in order to avoid any possible "paradoxical tension" and to make it practicable. One method of meeting criterion (2) is by integrating the principle of agent-relativity into expectabilism. Agent-relativity, according to Quong, is the recognition

> that each person is understood to have a powerful agent-relative permission to avoid sacrificing or significantly risking their own life for the sake of others (absent any obligations voluntarily incurred). (516-517)

By merging Quong's agent-relativity and Parfit's expectabilism, we create a theory of moral permissibility that may qualify as subjective rightness. It is an evidence-relative theory that recognizes self-preservation permissions. Agent-relative expectabilism might be defined as a theory of moral permissibility that holds:

Given no volition or voluntary obligations otherwise, it is permissible for a person to act in the way whose outcome would be expectably-best without risking self-sacrifice or significant risk to him or herself.

The agent-relativity principle incorporated in the agent-relative expectabilism above allows for self-preservation. It directly answers Question (b): what is the value of those unknown number of civilian and enemy combatant lives compared to the lives of the Marines? By stipulating that the Marines have a special interest in their own survival, by their own view, the value of their lives exceeds that of the enemy combatants and civilians. This stipulation will always provide guidance for actions, instead of only an after-the-fact means of evaluation. Because agent-relative permissions are a kind of freedom, agent-relativity allows for the moral permissibility of actions rather than moral imperatives that require actions.

Agent-relative expectabilism can achieve the ultimate goal of the theories of permissibility and international law listed above. This view can provide a framework for moral decision-making even when subjects are faced with uncertainty. It can be used by real subjects, who are bound by the constraints the real world places on knowledge. People can apply it with limited factual knowledge to whatever their situation may be, seeking guidance about the morality of their choices. Agent-relative expectabilism can provide that guidance. However, this view still has not satisfied criterion (3), to answer or provide a method of avoiding the problems posed by Question (a): what is the number of civilian and enemy combatant lives at stake? As a result, it has no real proportionality requirement.

Because of the phrasing of agent-relative expectabilism up to this point, there are no limits to the agent-relative permissions. Quong notes that a "difficulty with the agent-centered approach … is that it does not appear to draw … [a] distinction between Threats and Aggressors on the one hand and Bystanders on the other" (Quong 518). For the Marines, it will always be expectably-best that they survive, and the agent-relativity principle grants them the permission to make that a reality. As the theory currently reads, when applied to Combatants in a Cornfield, it would tell the Marines returning fire and killing both the enemy and any civilians is always permissible, whether there were

One enemy combatant and zero civilians, or
Ten enemy combatants and ten civilians, or
One enemy combatant and 20 civilians.

This is unintuitive, and it shows that agent-relative expectabilism is still incomplete. It cannot be permissible to return fire and kill 20 civilians in order to kill only one enemy. The third and final criterion of an effective action-guiding moral permissibility theory must be met because that will be the limiting aspect. But Question (a) is unable to be answered because the fact of the matter is unavailable. The purpose of Criterion (3) is to answer or provide a method of avoiding the problems posed by Question (a). If the question cannot be answered, then a method for avoiding the problems of proportionality it poses must be identified.

To satisfy the third criterion, the final aspect of the effective permissibility theory is to ensure that agent-relative permissions are not excessive but proportional to both the threat and the expectably-best outcome. Because the entire theory thus far has remained evidence-relative, the standard for that proportionality must also be evidence-relative. It should require "reasonable estimation" based on the evidence available to the subject. Evaluation of an act's expectable goodness requires an estimation of the goodness of its effects and an estimation of those effects occurring. The standard for proportionality should likewise require a reasonable estimation of the noncombatants who would be directly affected. That standard can be determined in a way very similar to "expectable goodness" and can be called

Expectable-noncombatant-cost (ENC): the reasonable estimation, using only the evidence available to the subject, of noncombatants who would be directly affected (both significantly harmed and killed) due to the act in question if that act were to occur.

It is unintuitive that one person can permissibly kill two noncombatants in order to save their own life. Any confusion on that point is due to scale. When groups are considered, such as in the context of war, wherein one squad of 15 could possibly save themselves by killing 30 noncombatants, any ambiguity or illusions are cleared up. 15 lives for 30 is clearly unintuitive and impermissible, and thus so is a single life for two. However, Quong claims that it can be permissible for one person to preserve their own life by killing one person who is "innocent" and who has "done nothing to make themselves liable to be killed" (Quong 532). He does not apply that thinking to groups. However, according to his analysis, a one-for-one exchange can be permissible.

Suppose the one-for-one exchange could apply to groups, or sides of a conflict. In that case, if the ENC of an act were to exceed that act's number of subjects (those appealing for permissibility-guidance), then the act would be impermissible. If the ENC were less than or equal to the number of subjects who would perform that act, then the act would be permissible. This evaluation can

be called the "ENC threshold", which is equal to the number of actors of an act, plus one. If the ENC threshold is not met, then the agents of the act in question retain their agent-relative permissions. If the ENC threshold is met, the agents lose those permissions. For example, in Combatants in a Cornfield, if the squad was composed of 15 Marines, then the ENC threshold would be 16 noncombatants. Were the Marines to reasonably estimate that 15 or fewer noncombatants would potentially be killed or significantly harmed, they would retain their agent-relative permissions. If they estimated 16 or more, they would lose their agent-relative permissions.

An effective action-guiding moral permissibility theory, then, must satisfy three criteria.(1) It must be be evidence-relative so it is practical for use in real situations. Expectabilism satisfies this criterion. (2) It must assess the value of the lives of both the subjects and the unknown numbers of enemy and civilians. The principle of agent-relativity satisfies this criterion. Finally, (3) it must have evidence-relative proportionality limitations for the permissions created by Criterion (2). The ENC threshold satisfies the third criterion.

To incorporate the ENC threshold into agent-relative expectabilism is the final task. The theory of ENC threshold agent-relative expectabilism is:

It is permissible for a person to act in the way with the expectably-best outcome, without risking self-sacrifice or significant risk to themselves (given no volition or voluntary obligations otherwise) if and only if the ENC threshold is not met.

For the Marines in the Combatants in a Cornfield scenario, this theory provides actionable permissibility-guidance. By incorporating a limited agent-relativity, the expectabilism problems in which "we are pulled to evaluate an agent's action as paradoxically both right and wrong" (Smith 67) are eliminated—concurrent and opposite answers no longer occur. Further, the ENC threshold places limitations on the agent-relative aspect of the theory, effectively making a reasonable proportionality stipulation. The ENC threshold agent-relative expectabilist analysis in Combatants in a Cornfield made by the squad leader or each Marine in the squad would consist of the following thoughts:

Expectably-best outcome: all Marines survive.

Volition or voluntary obligations to not survive: none.

Risk of self-sacrifice, significant risk to selves: not returning fire to eliminate any enemy.

ENC threshold: squad-

size population: extremely unreasonable and unlikely.

Returning fire and eliminating any enemy in the area is morally permissible.

The Marines are in the cornfield, facing a violent enemy, impending death, and uncertainty. But that very lifelike situation should not preclude them from being able to act, nor should it preclude them from being able to analyze the moral permissibility of different potential actions.

Restricted knowledge, a limited ability to gauge the truth of apparent facts, and uncertainty are common in the real world. They are so commonplace in war that the phrase "the fog of war" has come to be a direct reference to them. It is important to recognize that limitations in understanding the facts of a situation directly influence decision-making. It is also important to recognize that combatant participants in war are not necessarily and intrinsically immoral. Rather, in spite of the difficult, nasty, and possibly violent situations which they find themselves in, they may be trying to do what is right. Therefore, an effective moral permissibility theory needs to be able to not only evaluate actions in the past, but also guide actions in the present. An action-guiding permissibility theory needs to have properties that enable it to be applied in spite of the limitations mentioned above. Agent-relative expectabilism

with ENC threshold limitations may be a kind of permissibility theory that can be effectively used to guide actions.

REFERENCES

Haque, Adil Ahmad. "Killing in the Fog of War." *Southern California Law Review* 86.63 (2012): 63-116. Print.

International Committee of the Red Cross. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*, Eds. Yves Sandoz et al. Geneva: Martinus Nijhoff, 1987. Print.

---. "War and Law." ICRC. Web. 12 November 2014. < https://www.icrc.org/en/war-and-law>. McMahan, Jeff. "The Basis of Moral Liability to Defensive Killing." *Philosophical Issues* 15.1 (2005): 386-405. *Wiley Online Library*. Web. 12 November 2014.

---. "Torture in Principle and in Practice." *Public Affairs Quarterly* 22.2 (2008): 91-108. *JSTOR*. Web. 12 November 2014.

Parfit, Derek. *On What Matters: Volume 1* (The Berkeley Tanner Lectures). Oxford: Oxford UP, 2011. Print.

Quong, Jonathan. "Killing in Self

Defense." *Ethics* 119.3 (2009): 507-537. Print.

Smith, Holly M. "Subjective Rightness." *Social Philosophy & Policy* 27.2 (2010): 64-110. Print.

Thomson, Judith J. "Self Defense." *Philosophy & Public Affairs* 20.4 (1991): 283-310. Print.

Walzer, Michael. "Terrorism and Just War." *Philosophia* 34:1 (2006): 3-12. *SpringerLink*. Web. 12 November 2014.

NOTES

1) McMahan might classify those noncombatant civilians as bystanders or innocent threats. In either case, he would argue that it is impermissible to kill them.

2) While Smith devotes her entire essay to the topic, McMahan only touches on it, while many others examine only objective permissibility.

3) Parfit attempts to answer these questions and solve the problem they pose by discussing the relationship between knowledge, reasoning, and action. In his discussion, he defines different senses of the term "ought practically". Those senses are fact-relative, evidence-relative, belief-relative, and normative-belief-relative reasons to act a certain way. All of the senses of the term "ought practically" may, at least sometimes, be able to be disregarded as an extraneous complication.

FREE WILL AND TRANSWORLD IDENTITY IN THE METAPHYSICS OF LEIBNIZ

Jorge Aurelio Menéndez
Johns Hopkins University, Class of 2015

ABSTRACT

In his work in metaphysics, Leibniz argues for the reality of free will while rejecting the possibility of transworld identity. Here, I explore these two philosophical positions while showing that they stand at logical odds. In sections 1 and 2 I outline his argument for each of these points while showing that they both follow from the central tenet of Leibniz's metaphysics: that God is good. In section 3 I go on to demonstrate that they are in fact incompatible, under three different interpretations of free will found in Leibniz's work. In section 4 I argue that no matter how free will is interpreted, it seems that by definition it is incompatible with transworld identity under the Leibnizian conception of possible worlds. This turns out to be a significant source of tension within Leinbiz's metaphysics. I superficially explore one possible solution but leave this as an open avenue for further investigation.

LEIBNIZ THE COMPATIBILIST

Throughout the continuum of views expressed in Leibniz's work on metaphysics, there seems to be one common thread that he maintains very close to his mind and heart: that God is good. God is the great creator, responsible for everything that is, has been, and will be, and he creates the world with perfect harmony in mind. Leibniz clings on to this idea from day one, and it seems to be the seed from which many of his ideas grow.

One of the major implications of this notion is the fact that God must be able to choose freely when he makes decisions on how the world is. For him to be good in virtue of having actualized the best world possible for us, he must have had a choice between other possible worlds. He writes,

> (1) For God chooses among the possibles, and for that very reason he chooses freely, and is not necessitated; there would be neither choice nor freedom if there were but one choice possible. (*Theodicy* 235)

A major result of this, as hinted at by the quote, is a rejection of necessitarianism. While God must have chosen the actual world by virtue of his goodness, it remains true that there are several possible worlds from which he had to pick. This is the set of all internally consistent worlds, some of which are better than others and one of which – the actual world – is the most harmonious and hence best world. The fact that God chose to actualize this one, as opposed to the infinitude of other worlds he also *could* have actualized, makes him good.

But if God is good, why did he choose to actualize a world in which the Thirty Years War happened? To address this issue, Leibniz brings in free will to the picture. In his paper on "Necessary and Contingent Truths", he addresses this by making a distinction between God's decision to put a certain mind into the world – e.g. Judas – and the inherent nature of that mind:

> (2) … God first considers a mind as possible before he decrees that it should actually exist. For the possibility or notion of a created mind does not involve existence. … God, therefore, does not decree that Judas must be a traitor. All that he decrees is that Judas, whom he foresees will be a traitor, must nevertheless exist, since with his infinite wisdom he sees that this evil will be counterbalanced by an immense gain in the greater goods, nor can things be better in any way … . [I]t is the very notion of the creature, in so far as it involves limitation (which is the one thing that it does not have from God) that drags the act towards badness. ("Necessary and Contingent Truths" 104-105)

The most perfect and harmonious world is the one that contains Judas. So God chose it to be actualized. But he did not will for Judas to commit the evil acts he did – these acts were willed by the mind of Judas. Whereas God willed the existence of Judas's mind, Judas's mind was what willed the evil. Similarly, by actualizing the best possible world, God had to actualize the minds of several individuals who in turn were such that their behavior provoked the Thirty Years War. In actualizing the best world possible, God was forced to actualize certain entities that themselves willed unfortunate events to unfold within it. These entities thus in some sense have free will.

Leibniz argues for the reality of free will in many other texts, maintaining that indeed minds can make free choices, much like God does in choosing the world to actualize. Robert Adams writes in his book *Leibniz: Determinist, Theist, Idealist*, "Leibniz was a compatibilist, maintaining to the end of his life that every event is determined but some acts are nonetheless free" (Adams 11). But this turns out to be a significant source of tension within his metaphysics. In this paper I will investigate the tension generated by Leibniz's belief in free will, particularly with respect to his rejection of transworld identity.

LEIBNIZ ON TRANSWORLD
IDENTITY

Let us return to the Judas example. By choosing to actualize the actual world in which Judas betrayed Jesus, God seems to have at least indirectly caused evil in the world. Even if the very act was a product of the mind of Judas *per se* and not a direct act of God, God still seems to have created evil in the world in some sense. Leibniz cannot allow this. His central thesis revolves around God's goodness, so he must find a way to make it such that God is not responsible for Judas's evil actions, directly or indirectly. He does this by way of showing that God in fact had no choice but to actualize these evil minds because of the nature of the set of possible worlds he has to choose from. The construction of these worlds in turn ultimately leads to a rejection of transworld identity.

The first step in this argument is showing that God's choice in worlds is in fact quite constrained. Leibniz classifies the concept of a world as either *basic* or *complete*. The basic concept of a world contains everything about it that happens *within* it. Its *complete* concept additionally includes information regarding its relation to God's decision, i.e., whether it is the best world or not. All that Leibniz requires for his metaphysics is that the basic concepts of all worlds be internally consistent. This establishes their modal nature as possibilities, from

which follows the reality of God's choice (i.e., rejection of necessitarianism) and hence his goodness. While all worlds except for one will not be actualized (i.e., their complete concepts include that they are not the best world and hence won't be chosen by God), they still remain "possible in [their] own nature, even if [they are] not possible in respect to the divine will" (*Textes Inédits* 289). This is precisely what makes God's choice a free choice.

But for a basic concept to be internally consistent is not so simple. Every single element in the world and everything that happens to each one must fit together logically. It then follows that one can derive the entire structure of a world from just one of its elements. Leibniz writes,

> (3) it follows that the nature
> of every substance is such that
> by its force of acting and being
> acted on, that is by the series
> of its immanent operations, it
> expresses the whole universe.
> (*Die Philosophischen* VII 316)

So when God at the beginning of all time considers the plurality of worlds before him, this set, while infinite, is in fact quite constrained. These worlds are not random combinations of concepts; each one is a systematic combination of concepts that fit together in a logically coherent puzzle.

What this ultimately means is that no world can be exactly the same as

the actual one but without Judas's evil deeds because a non-evil Judas would not logically fit in with the rest of the world. If there were some world w that contained everything from the actual world except for Judas's evil deeds, it would then follow logically from the structure of w that Judas's evil deeds are also contained in it. If they were not, it would not be logically consistent. And, *vice versa,* it follows logically from the concept of a non-evil Judas that the actual world would be different. A Judas that did not commit any evil deeds – or even a Judas who avoided only one of his actual evil deeds – in fact bears no relation to the actual Judas, for it follows logically from his concept that the world around him would be completely different from the actual world. Note that the concept of non-evil Judas not only implies differences in the world after his existence (e.g., Jesus does not die and resurrect so Christianity never exists) but it also implies differences in that world pre-Judas, i.e., events that are logically required for non-evil Judas to exist (e.g., Judas's mother did not smoke). Importantly, these differences logically permeate down to the position of the nth atom at the moment of the Big Bang. Naturally, these differences would result in a world totally different from ours.

But then what is Judas? If non-evil Judas is allowed to exist, the concept of Judas must not include what evil Judas has done – this aspect must be flexible to allow the exclusion of

his evil deeds so that such a thing as 'non-evil Judas' may exist. But once we do this, it becomes impossible to derive the structure of the entire world from the concept of Judas, which Leibniz contends must be possible in (3). Because of this constraint on the construction of possible worlds – i.e., that the concept of an individual defines the world around it – an individual inherently cannot be part of any world other than the one it defines. Non-evil Judas is no longer Judas because he defines a completely different world.

From here follows rejection of transworld identity. For Judas can exist in one world only. Leibniz writes,

> (4) It seems to us indeed that
> this block of marble brought
> from Genoa would have been
> exactly the same if it had been
> left there, because our senses
> make us judge only superficially,
> but at bottom because of the
> connection of things the whole
> universe with all its parts would
> be entirely different, and would
> have been another [universe]
> from the beginning, if the least
> thing in it went otherwise than it
> does. (*Die Philosophischen* II 42)

Because the concept of an individual bears a certain relationship to the world around it – in that the full specification of that world is derivable from it – an individual inherently cannot be part of any other world than the one it defines. Transworld

identity is hence impossible.

In summarizing the present discussion, we have seen that it is central to Leibniz to have the notion of the "basic concept" of a world in order to defend God from any blame for the presence of evil in the world. This notion – that God's choice is constrained to the set of internally consistent worlds – allows Leibniz to show that any evils in the world follow from God's goodness in choosing the best world for us, for these evils turn out to be logical consequences of all the things that make the best world the best while keeping it logically consistent. However, it follows from this construction that the structure of any given world can be derived from the concept of any one of its individual members. And from here follows the rejection of transworld identity.

According to this interpretation of Leibniz's metaphysics, it is his central tenet of God's goodness that motivates his rejection of transworld identity. As I mentioned in §1, this principle is also what motivates his admission of free will. Unfortunately for Leibniz, however, it appears that these two ideas contradict each other. I will now turn to this contradiction.

THE CONFLICT

Leibniz's rejection of transworld identity and belief in free will, while motivated by the same thesis, in fact directly contradict each other. I will attempt to show this here.

Leibniz's notion of free will is somewhat vague, but I believe that every way in which he develops it seems to lead to a contradiction with transworld identity. Here I consider three readings that come through most strongly from Leibniz's work.

Free choice from lack of contradiction

The simplest notion of free choice that Leibniz works with is the idea that alternative choices do not imply a contradiction and hence are all possible. I could wear a red shirt tomorrow or I could wear a blue one. If I choose to wear a red one, this choice does not contradict anything else about what I choose tomorrow, and likewise if I choose to wear a blue one. Hence, both are possible choices, and I can choose freely between them. Leibniz writes,

> (5) [I]t is indeed always true that our *freedom*, and that of all other intelligent substances right up to God himself, is accompanied by a certain degree of indifference or contingency, which has been defined in such a way that we and those substances are never necessitated, since the contrary of that which happens always remains possible or implies no contradiction. (*Textes Inédits* 480)

This notion of free will is intimately related to the notion of the "basic concept" of a world outlined above. Even if the actual world is one in which I will choose to wear a red shirt, with regard to what happens in this world – without consideration of God's choice – it remains logically consistent for me to wear a blue shirt. My complete concept strictly includes wearing a red one (supposing I in fact do end up wearing a red shirt tomorrow), but at the local level of tomorrow's dressing moment without consideration of God's choice or the entire infinity of my complete concept, it is logically consistent for me to wear a blue shirt.

It seems we are already running into some problems here, for no two worlds can be exactly alike until one specific moment, as illustrated by the Judas and marble block examples discussed above. Since the structure of an entire world flows from the structure of any one of its individuals, any changes to that individual change everything else in the world, including that which existed prior and posterior to its own existence. So my putting on a blue shirt instead of a red one would be inconsistent with everything I have done.

Leibniz might argue against this by emphasizing that this alternative choice is still *locally* consistent if we do not consider any past events. But even at such a local level, my picking an alternative choice puts me in another world. A world is self-contained in that everything that happens in it is necessarily what happens in it. Within a given world, everything that happens is

necessary; this is the nature of each of the possible worlds that God considers when choosing which one to actualize. If this were not the case, there would be no way for God to know which world was best. Hence, at some time point t in a world where I exist, I either put on a red shirt or a blue shirt. I certainly could not have put on only a blue shirt and only a red shirt at the same time – this would be logically inconsistent. Hence, my putting on a blue shirt tomorrow morning is an event at time t in some world w and my putting on a red shirt tomorrow morning is an event at the corresponding time t' in some other world w'. But for this to be the case, I must exist in both w and w', which requires transworld identity! This notion of free will stands at logical odds with Leibniz's rejection of transworld identity. I cannot have the blue shirt alternative without existing in multiple worlds – free will seems to require transworld identity.

Free will from infinite analysis theory of contingency

A second notion of free will that Leibniz mentions in his writings springs from his infinite analysis notion of contingency. One of the ways in which Leibniz defines what it means for some aspect of the world to be contingent is as the lack of a finite proof to show why it is so:

> (6) [I]n necessary propositions one arrives, by an analysis continued to some point, at an

identical equation (and this very thing is to demonstrate a truth in geometrical rigor); but in contingent propositions the analysis proceeds to infinity by reasons of reasons, so that indeed one never has a full demonstration, although there is always, underneath, a reason for the truth, even if it is perfectly understood only by God. (*Textes Inédits* 303)

In other words, if a proposition cannot be proven to hold in a given world through a finite proof then it must be contingent.

We can then use this for a demonstration of free will. Leibniz claims in (5) that free will goes hand in hand with contingency, which can be demonstrated by the lack of a finite proof. In this vein, Robert Adams eloquently makes the connection, writing,

> (7) Our minds, being finite, cannot completely understand the motives of our choices, because they are infinitely complex. On the same ground, the connection between a free decision and its ultimate reasons or motives will be contingent, and cannot be demonstrated, in the senses laid down in the infinite analysis theory of contingency. (Adams 35)

Since a given choice we make cannot

be determined in a finite number of steps, it cannot be necessary. Hence the alternative choices are left open – they are contingent – and the choice is free.

But we again run into a similar problem as before. If we think about the alternative choices I could make about the color of the shirt I will choose to wear tomorrow morning, we are considering alternative choice scenarios in different worlds. My picking out a blue shirt or a red shirt is certainly contingent, but one choice leads to a scenario in which I wear a blue shirt and the other leads to a scenario in which I wear a red shirt. As discussed above, these two scenarios cannot logically exist in the same world. Each one must correspond to a different world. But that is impossible, for I do not exist in more than one world. The Jorge that wears the blue shirt cannot be the same Jorge. In this sense, then, the alternative choices in fact are not left open at all. Although which one is left open is not demonstrable by finite means, only one is actually possible. Therefore, my choice is not free.

Free will as a metaphysical force

One last way to concretize Leibniz's idea of free will that comes through in his writings is as a metaphysical force from within the free individual. This notion appears extensively in Leibniz's paper on "Necessary and Contingent Truths", although it is not as well developed as the other two flavors of free will discussed above. In the paper he writes,

> (8) But free and intelligent substances possess something greater and more marvelous, in a kind of imitation of God. For they are not bound by any certain subordinate laws of the universe, but act as it were by a private miracle, on the sole initiative of their own power, and by looking towards a final cause they interrupt the connexion and the course of the efficient causes that act on their will. ... For just as the course of nature is changed by the free will of God, so the course of the mind's thoughts is changed by its free will. (Necessary and Contingent Truths" 100)

There is something special about free beings that gives only them particular metaphysical forces that provide them with free will. The exact nature of these forces is entirely unclear, but the central idea is that free will is a metaphysical force.

Again, it is easy to see how this conception of free will does nothing to bypass the problem of transworld identity. While I may have the power to change "the course of the mind's thoughts" through my free will, it still remains impossible to have more than one alternative action available to me in a given moment, since each alternative can only occur in its own mutually exclusive world, and I can only exist in one. Hence, only the

alternative that exists in the actual world that I occupy is open to my choosing, and my free will is no longer free. That metaphysical force within me cannot, as Leibniz contends, change the course of the mind's thoughts. For if it did, I would have to be occupying a different world in which I could not be the same person. Or it would have to somehow change the structure of the actual world, which would lead to internal inconsistency, rendering this impossible. This strictly follows from rejection of transworld identity, again revealing an inconsistency with his endorsement of free will even as formulated thus.

A LEIBNIZIAN RESOLUTION?

I believe I have shown clearly that Leibniz's notion of free will is incompatible with his rejection of transworld identity. I have considered only those conceptions of free will that I myself have found in my interpretations of various of Leibniz's works, but I am willing to go so far as to claim that any other formulation of it will run into similar problems. Indeed, each of the arguments above are fundamentally the same.

Let me illustrate this point further. Any formulation of free will requires by definition that some agent in a given world be able to pick out any choice out of some set of possibilities. He or she must be able to carry out or choose any single member of that set. For example, let us define an abstract

set of possible actions in an arbitrary choice situation such that A and B are two of its members. For our agent to be free, he must be able to pick A and he must also be able to pick B within the same world. But Leibniz does not allow this because any given world is defined by its basic concept, which is, by definition, fully specified. If it were not, then God would not be able to know whether it was the best world. Thus, every world is constructed in such a way that it was pre-determined whether action A or action B was executed. Worlds are such that either only A occurs or only B occurs (and not both since this would be logically inconsistent). The Leibnizian world therefore is not open to free will.

Herein lies the fundamental problem: because an individual cannot have free will within a single world, he can only have free will across worlds. Hence, an individual must be able to exist across worlds to have free will: free will *requires* transworld identity. If an individual cannot exist across worlds, he cannot have free will in the transworld sense that Leibniz allows. Under Leibniz's formulation of possible worlds, free will is incompatible with a rejection of transworld identity.

For Leibniz to maintain the reality of free will along with his notion of possible worlds he must accept transworld identity. What does this imply? It is not so clear how far reaching the implications are, i.e., how much they would change

Leibniz's metaphysics. I leave this question mostly open for further investigation, but I will discuss one possibility: sacrificing the Conceptual Containment Theory of Truth.

The Conceptual Containment Theory of Truth (i.e., "the predicate is included in the subject") is the idea that the individual is defined by its "complete concept": everything that it is, has been, will be, and everything that it has done and will do. As we have seen, this definition is incompatible with free will. For free will to exist, alternative choices must be contained within the concept of the individual, requiring an entirely different definition. If Caesar had free will, then he could have chosen to cross the Rubicon or not to cross it at all. Whether he had or had not, he would still be Caesar. For this to be allowed, the "complete concept" of Caesar must be dropped and replaced by some other definition of the individual (e.g., David Lewis's counterpart theory). Of course, this has far-reaching implications. First, one would likely no longer be able to derive the structure of the entire world from the concept of a single individual. And, second and most important, transworld identity would be allowed, thus relieving the tension I have illustrated throughout this paper. It also follows that there could be two worlds exactly alike up to some point at which they diverge. One can easily see how these consequences resolve the problems outlined above.

One major consequence of this move

that Leibniz might strongly disagree with is that it seems to take away from God's infinite wisdom. For if at any point I can pick freely between putting on a red shirt or a blue shirt, how could God possibly know which one I would pick beforehand? If he predicted that I would wear a red shirt, it must still be possible for me to choose the blue one for free will to maintain its coherence. And if, at the outset of the universe, God did not know which one I would pick, how could he know that this world was the best one? I intuit that there are ways to resolve some of these problems, but I will leave them open as a further issue to examine.

CONCLUSIONS

It seems clear from the current investigation that Leibniz's assertion of free will is incompatible with his rejection of transworld identity. It does not seem to be due to a certain definition or specific claim of his, but rather a fundamentally logical contradiction present within the theory. One way I propose to solve this is by dropping the Conceptual Containment Theory of Truth, but this also happens to be one of the central tenets of Leibniz's philosophy. For free will to be a reality within Leibniz's metaphysics, sacrifices will have to be made. Unfortunately it is not as simple as accepting transworld identity, for this has far-reaching consequences. The tension that arises from this conflict is indeed deep. Resolving it is thus likely a complicated matter, if not impossible,

within the confines of what we know of Leibniz's philosophical beliefs. I leave this as a future direction of research.

REFERENCES

Works of Leibniz

Die philosophischen Schriften von Gottfried Wilhelm Leibniz. Ed. by C. I. Gerhardt. Berlin: Weidman, 1875–90. Reprint, Hildesheim: Georg Olms, 1965.

Necessary and Contingent Truths. c. 1686

Textes inédits. Ed. by Gaston Grua. Paris: Presses Universitaires de France, 1948.

Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil. 1710

Translations

All translations except for "Necessary and Contingent Truths" appear in:

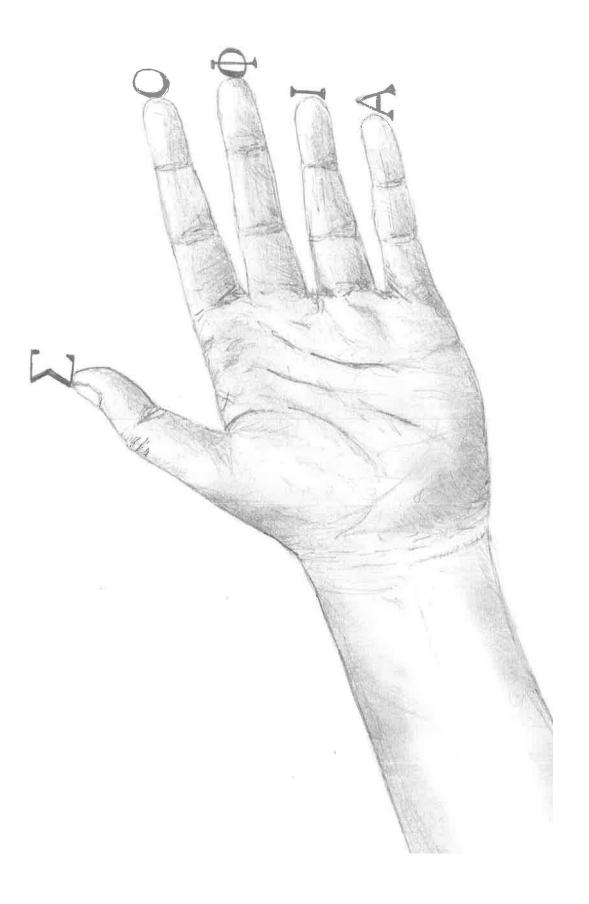Adams, Robert Merrihew.
*Leibniz: determinist, theist,
idealist.* Oxford University Press,
1994.

The translation for "Necessary
and Contingent Truths" appears
in:

Leibniz, Gottfried Wilhelm.
Philosophical Writings. Ed. G.H.
R. Parkinson. Totowa, NJ:
Rowman and Littlefield, 1975.

Secondary Sources

Adams, Robert. *Leibniz:
Determinist, Theist, Idealist.*
Oxford: Oxford University Press,
1994.

TWO KINDS OF ANTICIPATIONS IN PLATO: ANTICIPATIONS AS BELIEF AND IMAGINATION ABOUT THE FUTURE
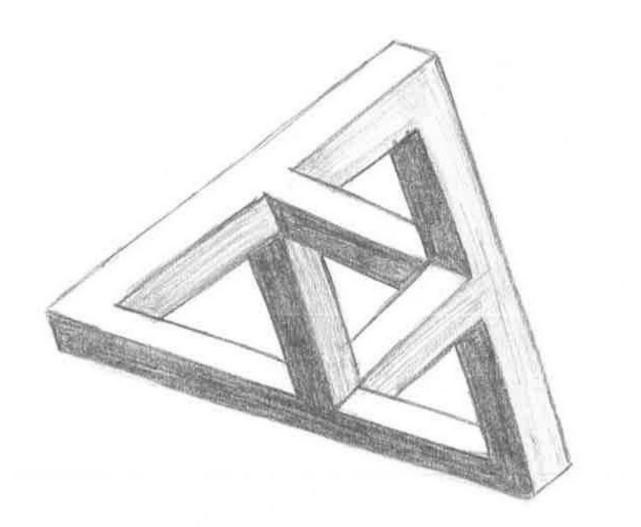
Wenjin Liu,
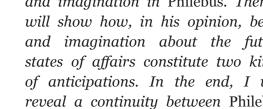*University of Pennsylvania, Class of 2015*

ABSTRACT

*In this paper, I am going to show how Plato's view about anticipation (*elpis*), a mental state that involves belief and imagination about the future, can offer useful insight into his claim in* Laws I *that pleasure (*hêdonê*), pain (*lupê*), daring (*tharros*), and fear (*phobos*) are four distinct non-rational motivating forces against the pull of calculation (*logismos*). To arrive at this point, I will first attempt to recapitulate Plato's account of belief and imagination in* Philebus. *Then, I will show how, in his opinion, belief and imagination about the future states of affairs constitute two kinds of anticipations. In the end, I will reveal a continuity between* Philebus *and* Laws. *Based on this connection, I will argue that by importing our interpretation of anticipation from* Philebus, *we can not only distinguish the four non-rational motivating forces from one another, but also explain how dar3ing and fear, anticipations or beliefs about future pleasure and pain, can lead us in directions different from the dictate of calculation.*

TEXT

Belief (*doxa*) is a controversial notion in Plato that has attracted an enormous amount of scholarly attention. On the one hand, many commentators believe that Plato's view about belief has undergone an obvious transition from the *Republic* to his later dialogues such as *Theatetus*, *Timaeus*, and *Philebus*. On the other hand, they disagree on the possible mental states that belief involves in later dialogues: it is controversial whether belief should be an exclusively rational and reflective mental state that can only be obtained by a conscious inner reflection. In comparison to the heated discussions that Plato's conception of belief has prompted, his view about the mental state that presents inner images (*eikonas* or *phatasmata*), imagination, has been studied less. Even fewer studies have been done on the relationship between belief and imagination. Such an omission not only restricts our understanding of belief, but also perpetuates our neglect of Plato's intriguing comment on anticipation (*elpis*)[1], a mental state that involves belief and imagination about future states of affairs.

In this paper, I am going to show how Plato's view about anticipation, a

mental state that involves belief and imagination about the future, can offer useful insight into his claim in *Laws* I that pleasure (*hêdonê*), pain (*lupê*), daring (*tharros*), and fear (*phobos*) are four distinct non-rational motivating forces against the pull of calculation (*logismos*). To arrive at this point, I will first attempt to recapitulate Plato's account of belief and imagination in *Philebus*. Then, I will show how, in his opinion, belief and imagination about future states of affairs constitute two kinds of anticipations. In the end, I will reveal a continuity between *Philebus* and *Laws*. Based on this connection, I will argue that by importing our interpretation of anticipation from *Philebus*, we can not only distinguish the four non-rational motivating forces from one another, but also explain how daring and fear, anticipations or beliefs about future pleasure and pain, can lead us in directions different from the dictate of calculation.

BELIEF AND IMAGINATION IN *PHILEBUS*

In this part of my paper, I will examine two kinds of mental states that Plato discusses in *Philebus*, belief and imagination, which can be called "anticipations" when they are about future states of affairs. In the first place, I will illustrate how Plato could offer a consistent account of belief-formation by examining the two relevant instances in *Philebus*. Then, I will look into Plato's conception of imagination, a mental state that

derives from belief. By analyzing his analogy of one's soul to a book and the metaphor of inner scribe and painter, I will reveal several essential features that Plato attributes to imagination.

To show that belief and imagination about the future are really anticipations, Plato begins with an examination of the formation and nature of belief (*doxa*) in *Philebus*. He demonstrates how we come to a belief or an attempt to come to a definite belief (to *diadoxazein egcheireîn*) by investigating how a man forms a belief about an object that he cannot see quite clearly.

S: Wouldn't you say that it often happens that someone who cannot get a clear view because he is looking from a distance wants to make up his mind about what he sees?
P: I would say so.
S: And might he then not again raise another question for himself?
P: What question?
S: What could that be that appears (*phantazomenon*) to stand near that rock under a tree? Do you find it plausible that someone might say these words when he sets his eyes on such appearances (*phantasthenta*)?
P: Certainly.
S: And might he not afterwards, as an answer to his own question, say to himself, " it is a

man", and in so speaking, would get it right?
P: No doubt.
S: But he might be mistaken and say that what he sees is a statue, the work of some herdsman?
P: Very likely.
S: But if he were in company, he might actually say out loud to his companion what he had told himself, and so what we earlier called *belief* (*doxan*) would turn into an *assertion* (*logos*)?
P: To be sure.
S: Whereas if he is alone, he entertains this thought (*dianooumenons*) by himself, and sometimes he may even resume his way for quite a long time with the thought in his mind?
P: No doubt.
S: But look, do you share my view on this?
P: What view?
S: That our soul in such a situation is comparable to a book? (*Phileb*. 38c4-e10; my italics)

According to this account of belief-formation, the first step of coming to a definite belief is being aware of some appearance. The man looks from a distance, so he cannot see clearly what the specific thing that appears dimly in his sight is. He only perceives what appears to him: something appears to stand near the rock under a tree (*Phileb*. 38c4-d1). However, neither the appearance presented to him

by perception nor the background information about the object can tell him what the object in question really is. Different from a case in which we can infer that a distant building must be a coffee house because we see people leaving with cups of coffee, the man cannot tell what the thing actually is on the grounds that it stands near the rock under a tree. In this way, Plato invites us to imagine that we are under some circumstances in which perceptual evidence underdetermines the identity of the objects in question.

In spite of the constraints of our perception, we often still want to know what an object really is. Similarly, the man in Plato's example of blurred vision wants (*boulesthai*) to determine what he sees (*Phileb*. 38c6). Such a wish is a double-edged sword. As early as in the *Republic*, Plato warns of the danger of entertaining appearances. In *Book* X, he harshly attacks imitative poetry on the grounds that it indulges the non-rational part of the soul by presenting deceptive images. Whenever a person is subject to a perceptual appearance that P, the non-rational part of his soul (*doxazei*)—forms a belief (*doxa*) that P. Admittedly, Plato suggests that we could proceed from unstable and misleading appearances to the possession of a higher kind of belief by means of exercising the calculating or rational (*logistikon*) part of the soul. Such a move requires measurement, counting, and weighing, which are usually effortful (*Rep*. 602c-603a). In fact, if we could always calculate, weigh,

and count the appearances presented to us instead of being mislead by them, Plato would not argue so adamantly for the banishment of imitative art in the city. Given these challenges of climbing the steep epistemic ladder from mere appearances, the man in the example of the blurred vision is liable to mistaken beliefs.

Beginning with an uncertainty about the appearances from perception and a wish to determine what the appearances in question objectively are, we often go on to an internal question-and-answer session. For instance, the man in Plato's example of blurred vision asks himself what the thing that appears to stand near that rock under a tree could be. Obviously, the clues he gets from perception are insufficient to answer this question: the object in question stands at a distance, so it can only render a vague impression; in addition, all other related perceptual experiences, for example, that a thing appears to be near the rock under the tree, are not enough for one to infer what the thing really is. Since Plato claims that we come to have beliefs or attempts to form beliefs from perception and memory, memory must step in when we want to arrive at definite beliefs, but our perception is not sufficient for us to do so.

While Plato regards memory as another source of belief-formation, he does not think it fares any better than appearances in terms of reliability or clarity. In *Philebus*, he is only

committed to giving an overview of memory: memory is the preservation of perception (*Phileb*. 34a9). This brief description of memory is expounded in *Theaetetus*. In this dialogue, Plato compares our capacity for forming memory to blocks of wax, which can vary over different dimensions such as their respective sizes and degrees of purity. Just as we make impressions from seal rings, we form memory when we put our thoughts and perceptions on the wax and imprint them on it. Having been generated in this way, each piece of memory is stored as an image (*eidolon*), and it lasts as long as the image is not rubbed off from the wax (*Theaet*. 191d4-10).

Given such a description of memory, we might naturally wonder how thoughts can be stored as images in our memory. One possible answer is that Plato does not think images have to be purely visual. Throughout his works, Plato categorizes different things as images. On the one hand, in the attack against imitative art, he claims that such art is far removed from the truth because it touches only a small part of truth, and this part is only an image (*eidolon, Rep*. 598b5). Hence, Plato obviously thinks our visual representations of physical objects constitute one subtype of images. On the other hand, he is willing to allow representations that involve other sensory modalities to be called images. For instance, he notes that a statement (*logos*) in a sense is the vocal image of thought (*dianoias en phonê hosper eidôlon*) (*Theaet*. 208c5).

Thus, an image (*eidolon*) can also be a vocal representation of mental content. Although these instances of images differ in the sensory modalities with which they are communicated, they all *represent* certain physical objects or mental contents. Hence, for Plato, images do not have to be exclusively visual; the essential feature of images is that they are representations of the original physical objects or mental contents. With this interpretation of image, we can now see that by saying that memory about previous thoughts and perception is stored as an image, Plato means each piece of memory represents the content of a thought or a perceptual experience.

Although memory preserves and represents our thoughts and perception, its impressions are mere images (*eidola*). For Plato, both an image and an appearance (*phantasma*; or some form of the verb *phantazomai* or *phainomai*, such as *phantazomenon* and *phantasthenta*) belong to the same ontological category. In his attack against imitative art in Book X of the *Republic*, Plato claims that a painter imitates the appearance of the bed, which is an appearance or a mere image rather than the truth or reality (*aletheia*) (*Rep*.598b1-3). When being viewed from different perspectives, physical objects do not accordingly change, but their appearances are constantly changing. In this way, Plato argues that images and appearances are ontologically distinct from the reality.

Back to the case of blurred vision, our digression into Plato's conception of memory reveals that perception and memory, the two sources of belief formation in this scenario, are mere appearances or images. They depart from the reality, so neither of them can reliably tell us what the object in question is. Moreover, appearances involve not merely vague perceptual impressions that distant objects render on us. As we have seen in *Republic* X, Plato counts our perception of a bed as an appearance since it is susceptible to changes when being viewed from different angles. Thus, he suggests that all impressions that we can get from perception, regardless of their apparent differences in clarity and reliability, are in fact only appearances. These appearances are removed from reality, so none of them can infallibly represent the actual states of affairs. If all perceptual impressions are appearances, the example of blurred vision will be a general model of belief formation. In all cases, we come to have beliefs by first entertaining perceptual appearances and memories, neither of which can provide us with foolproof clues about the reality.

With the aid of perceptual appearances and memories, the man in Plato's example of blurred vision answers his own question about the actual object behind its appearance. He could say to himself that it is a man and happen to get it correct (*epituchôs*), but he might also be mistaken and proclaim that it is a statue, the work of some

herdsman (*Phileb*. 38d4-8). Thus, according to Plato, one forms a belief by assenting to a statement that it is P as an answer to his initial doubt about the identity of the appearance. After having determined what an object behind its vague appearance is, one can pronounce his belief to others, and he thus makes a statement (*logos*). He can also keep this thought (*dianooumenos*) with himself for a long time (*Phileb*. 38e1-7).

Although in the case of blurred vision, Plato demonstrates that belief can come about from memory and perception to the affirmed statement that it is P by means of an inner discussion, he is willing to allow beliefs to involve a wider range of mental states. We can have beliefs not only about what a thing really is, but also about whether we feel pleasure or pain.

> S: Moreover, due to lack of memory, it would be impossible for you to remember that you ever enjoyed yourself, and for any pleasure to survive from one moment to the next, since it would leave no memory. But, not possessing *right belief*, you would not *realize* you are enjoying yourself even while you do, and being unable to calculate (*logismoû*), you could not *figure out* (*logizesthai*) any future pleasure for yourself. (*Phileb*. 21c1-6; my italics)

In this passage, when Plato refutes Protarchus' view that a life full of pleasures without any knowledge, intelligence, calculation, and other mental capacities is good, he observes that if people had no true belief, they would not realize they were enjoying themselves at the time when they were enjoying (*me doxazein kairein kaironta*, *Phileb*. 21c3-5). Merely undergoing the physiological changes related to pleasure is not sufficient for one to feel pleased. In order for one to experience pleasure, he must possess the true belief about his present experience: he must judge (*doxazein*) that it is pleasant. On account of this belief, which correctly associates his physiological state with the conception of pleasure, one enjoys himself.

In this case, belief is an evaluation of one's hedonic responses.[2] Judging, the activity of forming a belief, happens at the same time when people are experiencing physiological changes associated with pleasure or pain. It is not immediately clear how such an instantaneous appraisal should fit into the model of belief-formation as an internal discussion. As long as we have already established the associations between physiological changes and our corresponding hedonic responses, we seldom have to pause and reflect on what we actually feel. Most of the time, our beliefs that current experience is pleasurable or painful, and that we are pleased or pained, accordingly come about almost automatically. For example, suppose we generally enjoy drinking coffee, and we feel pleased

now as we are drinking coffee. The move from our activity of drinking coffee to the belief that we are pleased by it does not require that we conduct an inner discussion as the man in the example of blurred vision does. Hence, beliefs about our feelings of pleasure or pain seldom involve reflective thinking or inner question-and-answer sessions. If the inner discussion that Plato describes in the example of blurred vision is a necessary component of the formation of beliefs, then his claim that our realization about pleasure and pain is some kind of belief will be self-contradictory.

However, it is hardly conceivable that Plato offers two conflicting accounts in the same dialogue. Hence, his claim that beliefs are necessary for the awareness of pleasure and pain (*Phileb*. 21c3-5) and his description later that beliefs are formed by an inner question-and-answer session (*Phileb*. 38c4-e7) must be compatible. As we have seen, Plato recognizes the existence of a subtype of beliefs that are realizations of pleasure or pain. And when we become aware of pleasure or pain, we often do not conduct an internal dialogue to determine what we actually feel. Hence, the internal question-and-answer session by which the man in the case of blurred vision comes to a belief is not the essential component in the genesis of beliefs. As a consequence, although Plato presents a model of belief-formation with the case of blurred vision, he does not intend all beliefs to be formed

in strict accordance with this model.

In fact, by comparing the two accounts of belief in *Philebus*, we find an obvious and consistent parallel between these two apparently diverse ways of forming beliefs. In both cases, people form beliefs by assenting to the association between perceptual or quasi-perceptual experiences and the memory about these experiences. In the case of blurred vision, one finds a parallel between the appearances in his sight and his previous perceptual experience of some object in his memory, and he affirms that they are of the same sort. Similarly, when one believes that he enjoys himself, he grasps that his current feelings and his past experience of enjoying himself are the same kind of affections, and he assents to this identity relation. Presumably, the assent does not have to be made reflectively or consciously. That is to say, for example, we do not have to knowingly approve that our current experience is pleasant in order to feel pleased. Instead, the assent can be made implicitly, and it only reveals itself when someone asks us about our feelings. If we unconsciously assent to the statement that we feel pleased, we will answer that we feel pleased to the question about our feelings. In contrast, if we experience similar physiological states without implicitly endorsing that we feel pleased, we will not confidently answer that we feel pleased when we are asked about our feelings.

In this way, I have shown that in

Philebus, Plato conceives of beliefs as a range of mental states that involve assenting to the associations between perceptual or quasi-perceptual experiences and the related memory about these experiences. The process of forming associations can be conducted either reflectively as if people were discussing with themselves or immediately as if the connections between the experiences and their related memory just come about by themselves. To further illustrate his view about belief and imagination, which is a mental state derived from belief, Plato compares the souls of people to books with both words inscribed by a scribe and pictures painted by a painter. At the beginning of this analogy, he claims that he is mainly interested in examining some long-term mental contents: people might carry beliefs with themselves for a long time, and it is the soul in such a state that resembles a book (*Phileb*. 38e5-11). As he makes it clear later, Plato thinks that not only beliefs but also mental images can be preserved in people's souls for an extended period. By comparing beliefs and imaginations to words and pictures, which are the contents of the book, Plato implies that people can possess beliefs and imaginations without their constantly attending to them. Just as a book cannot have all of its contents displayed at once, a soul does not have to always entertain all of its beliefs and imaginations in consciousness, and it is unable to do that.

Using the metaphor of two craftsmen imprinting different kinds of contents in the book, Plato explains the formation and essential features of imagination. After memory and perception inscribe beliefs in people's souls just as a scribe writes words in a book, mental images evolve from existing beliefs as if a painter follows the words and makes illustrations (*eikonas*, *Phileb*. 38b1-5). In other words, Plato believes that beliefs determine the main contents of imagination. Let us revisit the case of blurred vision and see how it might work with this metaphor of two inner craftsmen. When the inner scribe inscribes one's belief that it is a man in the book, the painter does not have the autonomy to draw any figure other than a man, and thus he draws a man accordingly. However, because the words "it is a man" do not specify the circumstances in which one forms such a belief, the painter could convey these messages by portraying the man in a certain way. For instance, the man can be painted at the further side of the page, suggesting that he stands at a distance. Thus, with the metaphor of the inner scribe and the painter, Plato implies that although beliefs determine the main contents of sensory imagination, the latter could bring forth sensory information that does not feature in the foreground of belief.

According to Plato, the way in which we become conscious of episodes of imagination resembles sense-perception. After mental images are

formed in accordance with existing beliefs, we can be aware of episodes of imagination by seeing (*hora*) the images inside us. This similarity between the vision—an instance of perception—and imagination suggests that just as we do not fully control what we perceive, we often cannot consciously determine the contents of our imagination. As the example of blurred vision tells us, although we might form different beliefs based on the same piece of perceptual evidence, we cannot readily revise the content of our perception without changing our position: we could judge that it is a man or it is a statue, but we are not able to modify the content of our perception, an appearance, without adjusting our position. Similarly, when we *see* inner images, we often cannot readily revise their contents without changing the beliefs from which these images are derived.[3]

Plato's portrayal of imagination as internal pictures does not suggest that these mental images have to be visual. I have already shown that he is willing to allow a variety of things to be called images. Images can be the appearance of a bed as well as the pronouncement of one's belief. Since we have no reason to think that the image of a sound would be visual, Plato must allow imagination to represent impressions from a variety of sensory modalities. Therefore, while imagination is rendered as the process of producing images in the soul, these images should not be

understood as exclusively visual: the image of the painter should therefore be understood metaphorically.

ANTICIPATIONS AS BELIEFS AND IMAGINATIONS ABOUT THE FUTURE

In the previous part of my paper, I have shown that Plato examines two types of mental states. On the one hand, beliefs involve a wide range of mental states in which the subjects assent to the association between perceptual or quasi-perceptual impressions and memory. On the other hand, by means of an analogy between a soul and a book and a metaphor of two inner craftsmen, Plato illustrates that imaginations are quasi-perceptual representations of sensory impressions constrained by existing beliefs. In this part of my paper, I am going to demonstrate that Plato thinks both beliefs and imaginations can be about the future, and beliefs and imaginations about the future are together called anticipations (*elpides*). I will also examine specific cases of each kind of anticipation and inquire into their relationships with pleasure and pain.

Having examined the genesis of beliefs and imaginations and their connections, Plato claims that these two types of mental states can be about not only the past and the present, but also the future.

> S: And are those writings and pictures which come to be in us,

as we said earlier concerned only with the past and the present, but not with future?
P: Decidedly with the future.
S: If you say decidedly, is it because all of them are really anticipations (*elpides*) for future times, and we are forever brimful of anticipations, throughout our lifetime?
P: Quite definitely (*Phileb*. 39d6-e6; my italics)

Presenting his position in a negative question, Socrates gets the affirmative answer that all the writings and pictures they talk about before can be concerned with past, present, and future. We have already seen that in the metaphor of two inner craftsmen, the words of the scribe refer to beliefs that we form by affirming the connection between perception and memory, and the pictures refer to the sensory imaginations that we create based on existing beliefs. In this way, Plato shows that it is agreed that beliefs and imaginations can be about future states of affairs. These beliefs and imaginations about the future, as Socrates and Protarchus again admit, are really anticipations of future states of affairs. In this way, Plato shows that there are two kinds of anticipations: on the one hand, some anticipations are beliefs about the future; on the other hand, imaginations about future states of affairs can also be called anticipations. For the sake of brevity and clarity, I will use abbreviations for these two types of anticipations in the rest of my paper, so we have:

B-Anticipation: One subtype of anticipation in Plato that is in fact belief about future states of affairs.

I-Anticipation: The other subtype of anticipation in Plato that is actually a mental image (*eikôn*) or an appearance, derived from existing beliefs, about future states of affairs.

For the first type of anticipation, B-anticipations, Plato does not offer any specific example to explain how we can come to have it. At the first glance, there is an apparent challenge about how his earlier account of belief formation can accommodate the progression from past or present perceptual and cognitive resources to beliefs about future states of affairs.[4] Since Plato does not give a separate story for the formation of belief about the future, he must assume that it can fit into his general model of belief-formation in *Philebus*, which he has already illustrated with the example of blurred vision and the metaphor of the inner scribe. Presumably, we can have a belief that it will rain soon. To form this belief, we begin with a perceptual appearance that the sky looks cloudy. Since our memory tells us that in similar conditions, rain will soon come, the impressions from perception and memory concur at this particular occasion. Either reflectively, as if we have conducted an internal discussion, or almost automatically, we affirm the association between the appearance of a cloudy sky and the memory that it is likely to rain when the sky is overcast. In this way, we come to believe that it is going to rain.

While Plato is not committed to giving a detailed account of B-anticipations, he offers a vivid example of I-anticipations. As he observes, someone often envisages himself in the possession of an enormous amount of gold and of a lot of pleasures as a consequence. And that one also sees, in this inner picture himself, that he is beside himself with delight (*Phileb*.40a7-10). In this example, the verb observe or see (*hora*) agrees with the verb that Plato uses earlier in his discussion about the way in which one becomes aware of his imagination: he sees (*hora*) the image he has formed inside himself (*Phileb*. 39c1). In addition, the contents that the man becomes conscious of are called "painted images" (*phantasmata ezographemena*) or "inner picture" (*enezographemenon*). These two words are just alternative ways of conveying a sense that is similar to that of images (*eikonas*) (39b8) and pictures (*zographemata*) (39d6), the words that he uses earlier to illustrate the contents of imagination. Hence, the example of envisaged wealth is undeniably an example of I-anticipations, which are in fact imaginations about the future.

According to Plato, the contents of I-anticipations can involve a variety of things. First of all, there can be quasi-perceptual representations in the contents of I-anticipations. As Plato describes with the example of envisaged wealth, one can see in his inner eyes that abundant gold comes to his possession (*kruson gigonomenon aphthonon*) (*Phileb*. 40a8). This imagined condition is one of the painted images (*phantasmata ezographemena*) or an inner picture (*enezographemenon*), so we become aware of it as if we are seeing it. In fact, as we have already seen, Plato allows representations of perceptual experience from different sensory modalities to be called images (*eidola*): not only is the appearance of the bed an image (*Rep*. 598b), but also the pronouncement of one's belief is counted as a vocal image (*Theat*. 208c5). Since the words "painted image" (*phantasmata ezographemena*) and "inner picture" (*enezographemenon*) are just alternative ways of referring to "image" and "appearance" (*phantasma*; or some form of the verb *phantazomai* or *phainomai*, such as *phantazomenon* and *phantasthenta*), we can infer that Plato must think that the contents of I-anticipations can also involve the representations of perceptual experience from various sensory modalities.

In addition, Plato holds that anticipated hedonic responses can be the contents of I-anticipations. As we have seen in the case of envisaged wealth, one sees with his inner eyes many pleasures as a consequence of his coming into

the possession of a large amount of gold (*Phileb*. 40a8-9). Hence, Plato obviously thinks that pleasures and pains in the future, which are expected hedonic responses, can be at least an episode of an imagined scenario. In addition to being part of a whole imagined condition, hedonic responses can constitute a piece of mental image by themselves. For instance, in the case of the envisaged wealth, the man also sees a picture in which he enjoys himself immensely (*kai de kai enezographemenon auto on eph' hauto kaironta sphodra kathora*) (*Phileb*. 40a 9-10). The existence of such a purely hedonic picture should not be a surprise to us. As we have seen before, Plato acknowledges that the realization that one takes pleasure or pain is a belief.[5] Admittedly, he does not describe the genesis of beliefs about future pleasures or pains. However, we have shown that his general account of belief formation applies to the formation of B-anticipations, which are in fact beliefs about the future. As a result, we have no reason to assume that Plato's account of the formation of beliefs about future hedonic responses, which are a subtype of B-anticipations, diverges from his general model of belief formation in *Philebus*. With the possession of beliefs about prospective hedonic responses, one can accordingly form mental images, just as if there is an inner painter who assiduously make illustrations of the words of beliefs.

In addition to being the contents of anticipations, pleasures and pains

can also be the affective responses to our entertaining subtypes of I-anticipations. After Plato has illustrated that expected pleasures and pains can constitute the contents of I-anticipations, which are in fact imaginations about the future states of affairs, he claims that for good people, these painted anticipated pleasures are usually true, but for wicked people, the anticipated pleasures that are painted in their minds are false. In spite of the divergent quality of these prospective pleasures, both the wicked and the good take pleasure in their anticipations of pleasures. While the wicked take pleasure in false pleasures, the good rejoice at true pleasures (*pseudesin ara hedonais ta polla hoi poneroi kairousin, hoi d' agathoi ton anthropoid alethesin*) (*Phileb*. 40c1-2). In other words, by imagining the representations of prospective hedonic responses, we simultaneously experience pleasure or pain as a consequence.

Not only do we have hedonic responses as a result of entertaining I-anticipations about pleasure and pain, but we can also experience other affective responses by having B-anticipations concerning pleasure and pain. Although Plato does not illustrate this situation with as much vividness as he does for the I-anticipations about pleasure and pain, we can infer from his discussion of anticipatory pleasure and pain that B-anticipations cause affective responses when they are about

prospective pleasure and pain. As Plato puts it, anticipation for pleasure is pleasant (*hedu*) and comforting (*tharraleon*), while the expectation of pain is frightening (*phoberon*) and painful (*algeinon*) (*Phileb*.32c1-2). It is now clear that one kind of anticipation, I-anticipations, can be pleasant or painful when they regard prospective pleasure or pain. And we have shown that when we entertain I-anticipations that involve expected pleasures, we can feel pleased or pained as a consequence. Hence, we can infer that when the other kind of anticipation, B-anticipations, are about future pleasure or pain, they are accordingly comforting or fearful. And just as pleasant or painful I-anticipations make us feel pleased or pained, comforting or frightening B-anticipations should make us comforted or frightened as a result.

"Comforting" (*tharraleon*) and "frightening" (*phoberon*), the two adjectives that Plato uses to describe the affective aspects of B-anticipations about prospective pleasures and pains, foreshadow his view about the contrast between non-rational and rational motivating forces in *Laws* I. In one passage, he claims that pleasure, pain, and beliefs about future pleasures or pains, which together are called anticipations, are motivating forces different from calculation *(logismos)*. He further distinguishes two kinds of anticipations: belief about future pleasure is called daring (*tharros*), and belief about the future pain is fear (*phobos*). Now, it is clear that

"comforting" (*tharraleon*) and "frightening" (*phoberon*), the two adjectives that Plato uses to describe subtypes of B-anticipations about prospective pleasure and pain in *Philebus*, are cognates of the words daring (*tharros*) and fear (*phobos*) in *Laws*. In the next part of my paper, I will begin by inquiring into this correspondence between these two works in order to show how my interpretation of two kinds of anticipations in *Philebus* can shed fresh light on our understanding of *Laws* I.

ANTICIPATIONS ABOUT PLEASURE AND PAIN AS MOTIVATING FORCES DIFFERENT FROM CALCULATION

In the previous parts of my paper, I have shown that according to Plato, beliefs are mental states that include an assent to the association between perceptual or quasi-perceptual experiences and the memory of these experiences; imaginations involve quasi-perceptual representations of sensory impressions on the basis of existing beliefs. Subtypes of belief and imagination, beliefs and imaginations about futures states of affairs, constitute anticipations. For the sake of clarity, we use the words "B-anticipation" and "I-anticipation" to stand for anticipations that are in fact beliefs and imaginations about the future. According to Plato, anticipations can have a variety of contents; notably, pleasures and pains can constitute a whole episode of I-anticipation and B-anticipation. In

addition, by entertaining anticipations about prospective pleasure and pain, we can experience a variety of affective responses. For instance, imagining prospective pleasure or pain makes us pleased or pained; believing that pleasure or pain will come to us makes us comforted or frightened. In this part of my paper, I will show how my interpretation of two kinds of anticipations helps elucidate the distinction among pleasure, pain, daring, and fear, four kinds of non-rational motivating forces, and the further distinction between daring and fear, on the one hand, which constitute anticipations about pleasure and pain in prospect, and calculation, on the other.

In *Laws* I, Plato notes that inside each of us, there are four different non-rational motivating forces against the power of calculation (*logismos*). First of all, we have pleasure and pain, which are opposite and witless advisors (*sumboulo enantio te kai aphrone*). In addition to these two, we have beliefs about the future (*doxas mellonton*); their common name is "anticipation" (*elpis*): anticipation for pain is fear (*phobos*), and anticipation for the opposite of fear, namely, pleasure, is daring (*tharros*). And against all these forces, we have calculation that determines which of these motivating forces suggest better or worse things (*Laws* I. 644c-d2).

It is obvious that the second pair of motivating forces, fear and daring,

alternatively called anticipations or beliefs about future pain and pleasure, correspond to Plato's characterization of B-anticipations in *Philebus*. In *Philebus*, Plato makes a distinction between two kinds of anticipation: on the one hand, anticipations can be beliefs about future states of affairs, which I call B-anticipations; on the other hand, anticipations can be imaginations or representations of sensory experiences based on existing beliefs, which are called I-anticipations. When pleasure or pain constitutes the content of these two kinds of anticipations, B-anticipations become comforting or frightening, while I-anticipations are pleasant or painful. Note that the adjectives that Plato uses to characterize the contents of a subgroup of the B-anticipations, namely, beliefs about future pleasure or pain in *Philebus*, are cognates of the names of the beliefs about the future pleasure or pain in *Laws* I: the adjectives "comforting" (*tharraleon*) and "frightening" (*phoberon*) are cognates of the words "daring" (*tharros*) and "fear" (*phobos*). Thus, two of the four non-rational motivating forces are subtypes of B-anticipations, namely, beliefs about prospective pleasure and pain.

By looking closely into Plato's explanation of fear, a B-anticipation about future pain, we can get a better sense of how our interpretation of B-anticipations elucidates the roles that fear and daring have as motivating forces. In a later passage in *Laws*

I, Plato distinguishes between two opposing kinds of fear. On the one hand, we could fear bad things when we expect they will happen to us (*prosdokontes genesthai*).[6] On the other hand, we could have fear about our reputation (*doxan*), believing that people will think ill of us if we do or say something that is not fine; this latter kind of fear is also called shame (*aiskunen*) (*Laws* I, 646e2-647a1). These two kinds of fears, the simple fear and shame, involve beliefs that differ in complexity. When we experience simple fears, we do not just fear bad things *per se*. Bad things that happen to strangers might invoke our compassion or empathy, but we do not normally fear these things. Instead, we fear bad things since we expect that (*prosdokontes*, literally, "believe ahead") bad things will happen to us. As for shame, it happens to us when we entertain a belief about the conditional relation between two events and an implicit assumption that the consequent event is a bad thing when it happens to us. To put it in another way, we experience shame when we believe that if the antecedent event, that we do or say something not fine, obtains, the consequent event, that people will think ill of us, will happen. And we assume that the consequent event, namely, a bad reputation, is a bad thing for us. Thus, we are not afraid of a reputation by itself; we are afraid of the ill reputation that we will get if we say or do something that is not fine, and we assume that an ill reputation is a bad thing to have. Therefore, both

kinds of fear involve beliefs that a bad thing will somehow happen to us, though these beliefs differ in their complexity. Since daring and fear are parallel affective responses, our inquiry into fear suggests that daring should similarly involve a belief that pleasurable things will come to us.

Although daring and fear are in fact beliefs that pleasure and pain will come to us, we do not have to consciously conduct an inner discussion in order to form these mental states. As we have seen in *Philebus*, Plato acknowledges that our awareness of pleasure and pain are beliefs, for he notes that right belief is necessary for us to realize (*doxazein*) that we are enjoying ourselves when we do (*Phileb*. 21c3-5). Such an immediate evaluation of our subjective feelings can hardly be as consciously reflective as the internal question-and-answer session that the man in the example of blurred vision conducts to affirm an account. In this way, we conclude that the essence of belief is the assenting, either immediately or reflectively, to the association between our perceptual or quasi-perceptual experiences and relevant memories. Thus, when we experience daring or fear, we do not have to conduct an internal discussion in order to believe that pleasure or pain will come to us. Instead, based on our perceptual experience and memory, we can come to assent to these beliefs either immediately or reflectively. And by having comforting or frightening anticipations, we feel comforted or frightened as a

result. In this way, daring and fear, B-anticipations about pleasure and pain, involve both affirmations to the statement that pleasant and painful things will happen to us and our natural affective responses to their contents.

The four non-rational motivating forces, pleasure, pain, daring, and fear, are compared to iron strings that are tied to a puppet and lead it in directions different from where the finer golden string, calculation, does. Some commentators argue that according to Plato's characterization, these mental states cannot be four distinct sources of motivation. As Susan Meyer puts it in her paper, it seems to these commentators that as long as we are attracted to the pursuit of pleasure or the avoidance of pain, these pleasures and pains must be prospective, and hence anticipated. However, if pleasure and pain as motivating forces must be anticipated, they appear to be the same as daring and fear, which are anticipations about future pleasure and pain. Even though people could claim that the experience of current pain involves mechanisms for recoil, which are arguably motivated by pain, it is not clear what movement can be elicited by the bare feelings of pleasure. Admittedly, pleasure might make us inclined to remain in a similar experience. Nevertheless, as long as we try to prolong or recreate the pleasurable experience, the pleasure must involve some anticipation about the future (Meyer 354). As a result, it turns out that Plato only

shows the existence of three non-rational forces of motivation instead of four: pain, daring, and fear; the latter two are anticipations or beliefs about prospective pleasure and pain.

This textual difficulty can be solved if we notice the continuity between *Philebus* and *Laws*. In both *Philebus* and *Laws*, Plato holds that daring and fear are subtypes of B-anticipations: they are beliefs about prospective pleasure and pain. As motivating forces, daring and fear involve both an assent to the statement that pleasurable and painful things will happen to us and our natural affective responses to their contents. In addition, we have seen that current pleasure by itself cannot prompt movement, so pleasure and pain as motivating forces must be about future states of affairs. Recall that in the second part of this paper, we have found that I-anticipations are imaginations about the future. In addition, an episode of I-anticipation can be made exclusively from anticipated pleasure or pain. For instance, the man in the example of envisaged wealth imagines a picture in which he is experiencing immense pleasure. By entertaining pleasure or pain in anticipation, we can feel pleased or pained at the same time. Hence, the pleasure and pain in *Laws* I could in fact be the I-anticipations of prospective pleasure and pain, which in turn prompt feelings of pleasure and pain. That is to say, we can be attracted by the representations of upcoming pleasure and feel pleased

as a response; we can also be averse to the representations of imminent pain and thus feel pained. In this way, we can distinguish the four non-rational motivating forces from one another. Pleasure, pain, daring, and fear are I-anticipations and B-anticipations about future pleasure and pain.

The differences among the four non-rational motivating forces become more obvious when we import our interpretation of Plato's conception of anticipations from *Philebus*. As we have shown, pleasure and pain as motivating forces could in fact be I-anticipations about future pleasure and pain, and by entertaining these I-anticipations, we feel pleased or pained at this time. When we entertain I-anticipations, we do not have to assume that they are real or possible. For instance, the man in the case of envisaged wealth could imagine the scenario in which he obtains great wealth, and he feels pleased as a result of entertaining this instance of I-anticipation. However, when he is asked whether he thinks his imagination will come true, he could give a negative answer, although he still enjoys entertaining it. Similarly, I-anticipations of pleasure and pain, as non-rational motivating forces, could naturally attract or repel us to act in one way or another. However, when we are asked about our reasons for acting in this way, we might be at a loss: we are simply attracted by or averse to the prospect of pleasure or pain, even though such prospects could be irrelevant to us. In contrast,

we experience daring and fear because we believe that pleasurable and painful things will come to us. Although we do not always form such beliefs consciously or reflectively, when we are asked about the reasons for our actions, we can cite these beliefs as our reasons.

By noting the continuity between *Philebus* and *Laws*, we also come to see more clearly how daring and fear can be among the motivating forces against the pull of calculation in *Laws* I. As we have seen, daring and fear contain both a conviction that pleasure and pain will come to us and our instinctively affective responses of feeling comforted and frightened. Throughout the whole experience of daring and fear, we can reflect on our perceptual evidence and memory to determine what will happen to us; alternatively, we can simply affirm the statement that something pleasant or painful will come our way. However, experience of daring or fear does not direct us to deliberate about what measure we should take.

In contrast, calculation as a rational motivating force involves an activity of figuring out the means to reap anticipated pleasure. Admittedly, commentators have not yet settled on a general interpretation of calculation.[7] Such disagreements in our understanding of Plato's general conception of calculation should not prevent us from pursuing a more modest project, that is, to see what makes calculation, a rational motivating force, distinct from the non-rational

ones in *Laws* I. In his characterization of calculation as a rational motivating force, Plato says that calculation is an assessment of which of the pleasures, pains, daring, and fear in us is better or worse (*Laws* I. 644d2). Since we have seen that pleasure, pain, daring, and fear are different natural propensities to pursue pleasure and to escape from pain in the future, calculation as the rational evaluation of these different natural affections should similarly be regarding pleasure and pain prospectively. This sense of calculation, that is, calculation as an activity of evaluating future pleasure and pain, also appears in *Philebus*. As Plato notes, the power of calculation is necessary for us to figure out (*logizesthai*) the pleasure we will enjoy in the future (*Phileb.* 21c3-4). Thus, calculation as a rational motivating force involves an activity of figuring out the means to reap anticipated pleasure. While daring and fear involve our beliefs that pleasure and pain will happen to us and the affective responses as a result of endorsing these beliefs, they do not direct us to deliberate about what we should do. Nevertheless, with calculation, we no longer merely expect that pleasure or pain will come to us, and we accordingly feel comforted or frightened. We are able to devise our own approaches to pleasure or avoidances of pain.

CONCLUSION

Up to this point, we have seen that in *Philebus*, Plato holds that belief (*doxa*) involves a wide range of mental states in which the subject affirms the association between perceptual or quasi-perceptual impressions and memories. Derived from and constrained by existing beliefs, imagination represents sensory impressions of various modalities. On top of his characterization of belief and imagination, Plato demonstrates that anticipations (*elpides*) consist of beliefs and imaginations about the future, which we call B-anticipation and I-anticipation. Both kinds of anticipation can have pleasure or pain in their contents. When I-anticipation involves expected pleasure or pain, it is pleasant or painful; by entertaining a pleasant or painful I-anticipation, we feel pleased or pained. Similarly, when B-anticipation involves the expectation of pleasure or pain, it becomes comforting or fearful; holding a comforting or fearful B-anticipation makes us comforted or frightened. Noticing the continuity in Plato's conception of anticipation between *Philebus* and *Laws*, we have found that pleasure, pain, daring, and fear, the four non-rational motivating forces in the latter dialogues, are in fact I-anticipations and B-anticipations about prospective pleasure and pain. In addition, by identifying daring and fear as B-anticipations about future pleasure and pain, we have revealed that they can lead us away from calculation because unlike the latter, daring and fear do not involve deliberation about the means to obtaining pleasure or avoiding pain.

REFERENCES

Bobonich, Christopher. *Plato's Utopia Recast: His Later Ethics and Politics*. New York: Oxford University Press, 2002.

Delcomminette, Sylvain. " False Pleasures, Appearance and Imagination in the Philebus." *Phronesis*, *Vol.* 48, no .3 (2003): 215-237.

Lorenz, Hendrik. *The Brute Within: Appetitive Desire in Plato and Aristotle*. New York: Oxford University Press. 2006.

Meyer, Susan Sauvé. " Pleasure, Pain, and 'Anticipation' in Plato's Laws, Book I." In *Presocratics and Plato: A Festschrift in Honor of Charles H. Kahn*, edited by R. Patterson, V. Karasmanis and A.Hermann. 349-366. Las Vegas, Nevada: Parmenides Publishing, 2012.

Moss, Jessica. "Plato's Appearance-Assent Account of Belief." *Proceedings of the AristotelianSociety, Vol.* 114 (2014): 213-238.

Plato. *Complete Works*, Edited by John M. Cooper. Indianapolis: Hackett Publishing Company,1997.

Plato. *Philebus*. Translated by Dorothea Frede. Indianapolis: Hackett Publishing Company, 1993.

Plato. *Timaeus*. Translated by Donald J. Zeyl. Indianapolis: Hackett Publishing Company, 2000.

Wilburn, Joshua. "Akrasia and Self-Rule in Plato's Laws," *Oxford Studies in Ancient Philosophy, vol.* 43 (2012): 25-53.

NOTES

1) Many translators and commentators use the word hope for *elpis*. However, such an interpretation is liable to confuse Plato's conception of *elpis* with our contemporary view of hope, which are apparently different. As I will show in later parts of this paper, for Plato, *elpis* can be a belief or an imagination about future states of affairs. In addition, the contents of *elpis* are in no way limited to the things or states of affairs that appeal to us. For instance, in *Laws*, Plato claims that the *elpis* that pain will come to us is fear (*phobos*). In contrast, the word hope in English often refers to the mental state that involves a desire for a state of affair and a belief that it is possible to be obtained. We do not normally hope for things

that we are averse to. Hence, I will not translate *elpis* as hope in this paper. Instead, I will use the word anticipation for *elpis*.

2) Plato believes that pleasure and pain are parallel but opposite processes arise together in the common kind (*Phileb.* 31c2-3). As Frede puts it, Plato thinks that pleasure and pain are not independent phenomena, but they occur in things that represent a harmonious mixture (Frede, 30. fn.2). On the one hand, pain arises when the harmony in the living organisms is disrupted. On the other hand, pleasure happens when the nature of living organisms is restored. This parallel between the occurrence of pleasure and pain suggests that we become aware of these hedonic responses in similar ways. Hence, although Plato only explicitly says that right belief is necessary for us to realize (*doxazein*) pleasure, he must intend that it is also necessary for us to be aware of pain.

3) Since Plato argues that imaginations are derived from existing beliefs, he could acknowledge that we can somehow modify the contents of our imaginations by changing the beliefs on which they rely. For instance, if we are asked to imagine an apple, we will

probably come up with a mental image of a red, round, and juicy fruit. If we are then asked to imagine a yellow apple, which is a less common but still a possible variety of apples, we can entertain a mental image of a yellow, round, and juicy fruit. In this case, we revise the contents of our imaginations because we change the beliefs on which our imaginations depend. By default, we assume the prototype of apples to be red, round, and juicy. Hence, when we are asked to imagine an apple, we immediately come up with a mental image of a prototype apple based on this belief. However, when we are later asked to imagine a yellow apple, the belief that some apples can be yellow reveals itself from our memory. Based on this different belief, we can accordingly come up with a mental image of a yellow, round, and juicy fruit.

4) In the next part of my paper, I will show how two subtypes of B-anticipation, daring (*tharros*) and fear (*phobos*), which are beliefs for future pleasure and pain, can play the role of motivating forces against calculation (*logismos*) along with pleasures and pains.

5) When Plato refutes Protarchus' view that a life full of pleasures without

any knowledge, intelligence, calculation, and mental capacities of such sort is good, he observes that if people had no true belief, they would not judge they were enjoying themselves at the time when they were enjoying(*me doxazein kairein kaironta*)(*Phileb.* 21c3-5).

6) Keen readers might notice that in the passages that we have examined before (*Phileb.* 32c1-2; *Laws* I. 644c-d2), Plato suggests that fear is a belief or anticipation about future pain. However, here, Plato tells us that we fear is a belief that bad things (*kaka*) will happen to us. This difference in Plato's choice of words implies that Plato might think that pain or painful things are in fact bad things that befall us. If this is the case, then he will hold that pleasure or pleasurable things are good things that happen to us. Due to the limit of space, I will not explore this issue further.

7) For instance, Bobonich construes calculation as practical deliberation that delivers all things considered judgments about what is good or bad for the person in the long run (Bobonich 2002, 263-7). In contrast, Wilburn proposes that calculation is a commitment to general principles of rational conduct, distinct from any

practical deliberations.

"Education is not preparation for life;
education is life itself."

-John Dewey