Pensées Canadiennes

Pensées Canadiennes

Pensées: The Canadian Undergraduate Journal of Philosophy
c/o McGill University Department of Philosophy
Room 908, Leacock Building
855 Sherbrooke St. W.
Montreal, QC
H3A 2T7

journal@penseescanadiennes.com
www.penseescanadiennes.com

CONTENTS /
TABLE DES MATIÈRES

INTRODUCTION

Welcome to the tenth edition of *Pensées Candiennes*. First published in 2002, *Pensées* is a national philosophy journal that showcases the work of talented undergraduate students to a wide audience, and in so doing promotes philosophy and philosophical discourse throughout Canada. Although the journal's editing team traditionally consists of philosophy undergraduates from McGill overseeing an editorial board of students from across Canada, this edition marks the first time that the position of editor-in-chief is shared by undergraduates from two different universities: McGill and the University of Montreal. We hope that this nascent partnership between philosophy students from the two schools continues well into the future, as it exemplifies *Pensées*' commitment to bilingualism and collaboration.

Inside this edition are four exceptional essays written by undergraduate students from four Canadian universities. Emma Ryman (McGill) writes on the phenomenological experience of auto-affection; Valérie Lynn Therrien (UMontreal) on the history of logic; Luke Kersten (UOttawa) presents a critique of the cognitive equivalence argument; and Edward Brook (Queen's) discusses intentionality in Brandom and Dennett. In this edition also appears an interview (in both the original French and its English translation) with distinguished Canadian philosopher Jean Grondin, in which we discuss truth, knowledge, and the future of philosophy.

We hope you enjoy reading.

Yours,

Lyndon Entwistle and Maxwell Ramstead
Editors-in-Chief, Spring 2012

INTRODUCTION

Bienvenue à cette dixième édition de *Pensées canadiennes*. Publiée depuis 2002, *Pensées* est un journal de philosophie national dont l'objectif est de souligner le travail accompli par des étudiants talentueux au premier cycle en philosophie. Par là, *Pensées* espère amener cette discipline à un large public et promouvoir le discours philosophique à travers le Canada. Cette année marque la première édition issue d'une collaboration interuniversitaire : les rédacteurs en chef proviennent de McGill et de l'Université de Montréal et supervisent, comme aux années précédentes, une équipe de rédaction provenant de partout au pays. Nous espérons que cette collaboration naissante entre étudiants en philosophie provenant de différents établissements se poursuivra, puisqu'elle exemplifie le dévouement de Pensée envers le bilinguisme et la collaboration.

La présente édition présente quatre essais remarquables écrits par des étudiants au premier cycle, provenant de quatre universités canadiennes. L'essai d'Emma Ryman (McGill) vise l'expérience phénoménologique de l'auto-affection ; celui de Valérie Lynn Therrien (Montréal), aborde l'histoire de la logique ; Luke Kersten (Ottawa) nous présente une critique de l'argument de l'équivalence cognitive ; et Edward Brook (Queen's) nous propose une étude de l'intentionnalité chez Brandom et Dennett. Cette édition présente aussi une entrevue avec l'important philosophe canadien Jean Grondin, où il est question de vérité, de connaissance, et de l'avenir de la philosophie.

Nous vous souhaitons une bonne lecture.

Bien à vous,

Lyndon Entwistle et Maxwell Ramstead
Rédacteurs en chef, printemps 2012

CONTRIBUTORS / CONTRIBUTEURS

Edward Brook recently graduated from McGill University, where he completed an honours degree in philosophy. His thesis, "Moral Codifiability: A Bifurcated Approach", explored the degree to which moral rules can be codified and categorized. He is interested in normative ethics, legal theory, and the philosophy of language. Edward is currently studying law at Queen's University where he hopes to one day combine his interest in language and ethics with his passion for legal theory. In particular, he would like to explore the roles that community sanctioning plays in the establishment and development of the criminal law system.

Edward Brook a récemment obtenu son baccalauréat en philosophie, dans le programme d'honneurs, à l'université McGill. Sa thèse, "Moral Codifiability: A Bifurcated Approach," a exploré les degrés auxquels les règles morales peuvent être codifiées et catégorisées. Il est intéressé par l'éthique normative, la théorie légale, et la philosophie du langage. Edward étudie présentement en droit à l'université Queen's, où il espère un jour pouvoir combiner son intérêt pour le langage et l'éthique avec sa passion pour la théorie légale. Il souhaiterait en particulier explorer les rôles que jouent les sanctions communautaires dans l'établissement et le développement du système pénal.

Luke Kersten is a senior undergraduate student at the University of Ottawa, specializing in philosophy. His primary philosophical interests reside in the philosophy of mind and moral psychology. He is fascinated by how neuropsychology and cognitive science can impact contemporary philosophical debates about the nature of the mind and morality. Luke is pursuing graduate studies in philosophy at the University of McMaster in the fall of 2012.

Luke Kersten termine son baccalauréat spécialisé en philosophie à

l'Université d'Ottawa. Ses intérêts philosophiques principaux sont pour la philosophie de l'esprit et la psychologie morale. Il est fasciné par l'impact que pourront avoir la neuropsychologie et la science cognitive sur les débats philosophiques contemporains concernant la nature de l'esprit et de la moralité. Luke poursuit ses études supérieures en philosophie à l'université McMaster en automne 2012.

Emma Ryman is an undergraduate student in the final year of her Honours Philosophy degree with minors in Classics and Political Theory at McGill University. She is primarily interested in moral and political philosophy, but also enjoys studying French and German phenomenology. Currently, Emma is working on her undergraduate thesis in normative ethics. Starting in the fall, she will be pursuing a Master's degree in philosophy at the University of Western Ontario. In her free time, Emma acts as one of the co-presidents of the McGill Philosophy Students' Association, and she also enjoys writing poetry and singing in her folk-pop band.

Emma Ryman termine sa dernière année au baccalauréat en philosophie, dans le programme d'honneurs, avec une mineure en classiques et en théorie politique, à l'université McGill. Ses intérêts principaux sont la philosophie morale et politique, mais elle aime aussi la phénoménologie française et allemande. Elle travaille présentement sa thèse de baccalauréat en éthique normative. L'automne prochain, Emma poursuivra ses études supérieures avec une maîtrise en philosophie à l'université de Western Ontario. Dans son temps libre, elle est co-présidente de l'association des étudiants en philosophie de McGill; elle aime aussi écrire de la poésie et chanter avec son groupe de musique folk-pop.

Valérie Lynn Therrien is currently an undergraduate student at The University of Montreal, and will be obtaining her Bachelors in Philosophy this year. She will be pursuing a Master's degree next year at the University of Western Ontario on a full scholarship. Valérie is currently planning a thesis that will focus on the contributions of the

Lwów-Warsaw School of Logic to modern logic through its remarkable programme aimed at solving classic philosophical problems with logic – most notably, Łukasiewicz's refutation of determinism through his innovative development of the multivalent propositional calculus. In her spare time, she works as a Unit Coordinator at the Montreal General Hospital and drinks a lot of coffee.

Valérie Lynn Therrien est étudiante en philosophie à l'université de Montréal, en voie d'obtenir son baccalauréat. L'an prochain, elle poursuivra ses études à la maîtrise à l'université de Western Ontario, où elle est boursière. Elle planifie actuellement une thèse qui se penchera sur les contributions à la logique moderne de l'école de Lwów-Warsaw, notamment le programme de recherche qui vise à solutionner les problèmes classiques de la philosophie en employant la logique; en particulier, elle voudrait travailler la réfutation du déterminisme de Łukasiewicz, à travers son développement d'un calcul propositionnel multivalent très innovateur. Dans son temps libre, elle travaille comme coordinatrice d'unité à l'Hôpital général de Montréal et boit beaucoup de café.

ENTREVUE
AVEC JEAN GRONDIN

Jean Grondin (né en 1955) est un philosophe canadien, professeur titulaire à l'Université de Montréal depuis 1991 et membre de la Société Royale du Canada depuis 1998. Il a effectué ses études en philosophie au premier et au second cycle à l'Université de Montréal; il a par la suite effectué sa thèse doctorale sur le concept de vérité en herméneutique à l'Université de Tübingen. Professeur Grondin est un spécialiste distingué de la pensée d'Emmanuel Kant, de Hans-Georg Gadamer et de Martin Heidegger. Ses recherches portent sur l'herméneutique, la phénoménologie, la philosophie classique allemande, l'histoire de la métaphysique et la philosophie de la religion. Nous avons eu l'occasion et le plaisir de nous entretenir avec lui.

PENSÉES: Qu'est-ce que la philosophie, pour vous ? Qu'est-ce qui vous a mené vers cette discipline ?

JEAN GRONDIN: Je vous remercie d'abord de cet entretien et d'autant que le dialogue est l'élément de la philosophie, qu'il s'agisse de celui que nous conduisons avec d'autres ou du dialogue incessant de l'âme avec elle-même. Ce qu'est la philosophie ? Je n'ai absolument rien contre sa définition classique: l'amour de la sagesse. Vous en connaissez une meilleure ? Ce qui m'y a mené ? Le fait que je ne l'aie pas, la sagesse, donc le souhait d'apprendre de ceux qui en ont ou plus que moi. On en vient à la philosophie parce que ses questions nous passionnent. Ce ne sont certainement pas les perspectives d'emploi qui nous y attirent. Il y a donc un risque pour tous ceux qui se lancent dans la philosophie. C'est une situation que l'on accepte parce que l'on ne peut pas faire autrement. L'homme est une question, une énigme pour lui-même, et on découvre des interlocuteurs aussi passionnants que rigoureux qui nous aident à poser cette question et, parfois, à esquisser des réponses.

PENSÉES: L'homme est une question ! C'est bien dit. Le dialogue, la question et la réponse sont effectivement au cœur de l'entreprise philosophique, et certains ont su y voir le point de départ réel, toujours déjà opérant, de la philosophie. Toute question, toute réponse se situent dans un dialogue qui les préexiste. Cet élément fondamental représente un aspect important de la pensée herméneutique, particulièrement celle de Hans-Georg Gadamer, que vous connaissez bien. Pouvez-vous nous dire quelques mots sur votre maître et sur le courant herméneutique, auquel vous vous êtes particulièrement intéressé?

GRONDIN: Bien dit, donc cela ne peut pas être de moi. C'est Augustin qui le dit au 10e livre de ses *Confessions*. Gadamer est effectivement un autre grand maître pour moi. Difficile cependant de n'en dire que "quelques mots" et cela vaudrait aussi du courant herméneutique. Que dire et par où commencer ? Gadamer est d'abord un philosophe formidable, que j'ai eu la chance de rencontrer et de bien connaître au fil des ans. C'est un pénétrant défenseur de l'humanisme, dont il nous rappelle d'ailleurs le sens au début de son chef-d'œuvre *Vérité et méthode*, dont la principale idée est que notre conception de la vérité et de la connaissance est beaucoup trop déterminée (sans que nous ne nous en rendions parfaitement compte) par la perspective un peu étroite de la science méthodique et de sa visée technique. Il est indéniable que la vérité est très souvent comprise à travers ce prisme, qui comporte bien sûr sa légitimité incontestable, car il incarne une voie privilégiée pour atteindre le vrai. Mais Gadamer soutient qu'il ne s'agit pas de la seule, qu'il y a une vérité de la sagesse pratique (ou de la prudence), de l'expérience acquise, de l'art, de l'histoire et de la philosophie elle-même. Il a raison de dire que le privilège du savoir méthodique tend à faire oublier ces expériences de vérité que *Vérité et méthode* nous aide à redécouvrir. Gadamer le fait en s'inspirant de la philosophie grecque, médiévale et moderne et de toute la tradition de l'herméneutique, ancienne et moderne, qu'il a portée à l'attention générale. L'herméneutique, pour faire court, est la discipline qui ré-

fléchit sur les grands principes de l'interprétation, notamment dans les disciplines qui y ont constamment affaire comme l'exégèse, le droit, l'histoire et la littérature. Mais c'était une discipline peu connue avant Gadamer. L'un de ses grands mérites est de l'avoir fait connaître et d'en avoir fait une philosophie incontournable.

PENSÉES: Cette remise en question de la notion même du vrai, comme absolument donné et objectif, semble aussi être au cœur de vos questionnements. Toute "vérité" est, de fait, traversée par les trames de l'interprétation, et ce même au sein des sciences dites "exactes", où, depuis au moins le début du XXe siècle, la notion de l'interprétation devient de plus en plus centrale. Il n'existe pas de faits "nus", pour ainsi dire : telle est l'une des leçons de l'herméneutique. À quoi attribuez-vous cette prise de conscience du caractère toujours herméneutique, sujet à interprétation, de la pensée philosophique ?

GRONDIN: Jamais, jamais je ne dirais qu'il s'agit d'une mise en question de la notion du vrai, bien au contraire. Où allez-vous chercher ça ? Je parlerais plutôt d'une redécouverte. La vérité reste pensée comme adéquation de l'intelligence au réel et pour moi il y a bel et bien des faits, nus ou habillés. Mais vous avez raison, parfois il faut effectivement un peu d'interprétation pour les trouver.

PENSÉES: La remise en question du vrai que j'ai suggérée ne signifie pas que la notion de vérité soit récusée, bien au contraire ; seulement, en mettant au jour les conditions de sa possibilité, il semblerait que l'on doive se faire à l'idée moins naïve, selon laquelle toute vérité est en partie dépendante d'un cadre d'interprétation et d'une méthode qui ouvre à son accès. C'est en ce sens que j'affirmais que la notion de vrai est remise en question par l'herméneutique. La voie d'accès au vrai est ainsi toujours issue d'un jeu d'interprétation ; elle sort d'un processus de questionnement qui doit lui-même être explicité, et non simplement tenu pour acquis. S'il y a bien des faits, et si la vérité est toujours pensée comme *adæquatio intellectus et rei*, il faut tout de même dire qu'après le tournant herméneutique, la notion

d'interprétation devient cruciale à la possibilité même du discours vrai. Je voulais en venir à l'interaction constante entre vérité, d'une part, et méthode, d'autre part.

GRONDIN: En quoi toute vérité est-elle dépendante d'un cadre d'interprétation ? Et faut-il à tout prix parler de "cadre" ? Qui le connaît au juste ? La meilleure interprétation n'est-elle pas celle qui ne se laisse limiter par aucun cadre et qui s'ouvre, tout simplement, à ce qu'il s'agit de comprendre ? Vous voyez, c'est moi qui pose des questions maintenant. Je suppose que cela est permis en philosophie. Quant au lien entre la vérité et la méthode, vous avez raison de parler d'interaction: c'est souvent la vérité qui commande la manière de la traiter. Aristote le disait déjà. En ce sens, l'herméneutique s'oppose à l'idée cartésienne d'une méthode universelle qui serait la condition de toute vérité. Tout champ aura sa méthode, tout objet ayant ses exigences auxquelles la pensée, et dès lors, son approche méthodique, doivent se conformer.

PENSÉES: Effectivement, s'il y a bien une discipline où le destinataire peut subitement devenir destinateur, c'est bien au sein de la philosophie. C'est bien le caractère dialogique de la philosophie qui lui confère toute sa force dialectique.

GRONDIN: Et à notre entretien tout son sens.

PENSÉES: Comme vous le dites, suivant Gadamer, les cadres interprétatifs ont eu pour effet de baliser notre conception de la vérité. Or, c'est bien cette conception pragmatique et technocentriste du savoir qui semble dominer à l'heure actuelle. En cela, serions-nous héritiers du projet moderne, cartésien ? Par ailleurs, croyez-vous que nous pourrions aller jusqu'à dire que la perspective herméneutique renverse le projet fondationnaliste de Descartes et son emphase, peut-être excessive, sur l'acquisition méthodique du savoir ?

GRONDIN: Dans les faits, Descartes est assez peu visé par Gadamer, directement du moins. C'est que la conception de la méthode que

se fait Descartes – une méthode déductive selon laquelle les connaissances doivent découler d'une certitude première – est bien différente de celle qui a pris le dessus dans la science moderne, méthode où la vérité repose moins (voire pas du tout) sur une certitude primitive que sur l'idée que les résultats de la connaissance doivent être indépendants de l'observateur et fondés sur l'observation. Cela marche en sciences pures, mais peut-on appliquer sans plus ce modèle aux sciences humaines, où l'implication du chercheur dans ce qu'il connaît n'est pas négligeable ? C'est cet apport productif du chercheur à ce qu'il comprend que Gadamer veut mettre en valeur dans sa conception de la compréhension comme fusion des horizons (de l'interprète et de la chose à interpréter). Vous avez raison de dire que c'est sans doute Descartes qui a ancré dans les esprits cette priorité de la méthode. Tout n'est cependant pas simple affaire de méthode, même chez Descartes, voire surtout chez lui. C'est que l'exigence de la méthode procède très clairement (si vous me passez l'expression...) d'un *ego* qui doute, qui hésite (*cogitare* veut d'ailleurs dire en latin classique "hésiter", c'est-à-dire "co-agiter" des possibilités), qui est donc en quête d'orientation, de sens et de vérité. La vérité méthodique présuppose ici une implication du penseur et, plus fondamentalement encore, la finitude. Il faut d'ailleurs se croire bien malin pour penser que l'on peut critiquer Descartes impunément sans prendre en compte l'ensemble de son corpus. Ce serait comme critiquer Dante ou Molière.

PENSÉES: Monsieur Grondin, je vous remercie sincèrement de nous avoir accordé cette entrevue. L'entretien aura été des plus intéressants. Nous voyons bien l'implication du penseur dans son questionnement, au sein de vos réponses ! Il est fascinant de voir à l'œuvre cette pensée du questionnement, ce questionnement de la pensée. Pour terminer, auriez-vous un message à transmettre aux philosophes de demain ?

GRONDIN: N'oubliez pas ceux d'hier. Ils vous ont ouvert la voie et permis de penser. Méfiez-vous de la bien-pensance du moment, des

querelles d'école et des chapelles sectaires, étrangères à la philosophie durable. Ne soyez pas trop déçus si la belle discipline de la philosophie est peu reconnue dans ce qui s'appelle le "vrai monde" (où sommeillent pourtant deux gigantesques notions philosophiques): comme l'a dit Descartes quand il a parlé de l'arbre de la philosophie, la philosophie essentielle, ou la métaphysique, s'interroge sur les racines de ce qui est, dont il va de soi qu'elles ne sont pas toujours visibles. Ne soyez pas trop découragés non plus si les échos (véritables, s'entend) à votre travail semblent modestes: les idées philosophiques mettent du temps à faire leur chemin. De toute façon, ce qui importe, ce n'est pas le succès immédiat, mais la pensée de demain, qui est entre vos mains. Je vous remercie de l'initiative de ce dialogue.

PENSÉES: C'est moi qui vous remercie, Monsieur Grondin, pour cette fusion des horizons!

INTERVIEW
WITH JEAN GRONDIN

Jean Grondin (born 1955) is a Canadian philosopher and professor at the University of Montreal, where he has taught since 1991; he was made a member of the Canadian Royal Society in 1998. Professor Grondin received a B.A. and M.A. in philosophy from the University of Montreal, and then completed his doctoral thesis on the concept of truth in hermeneutics at the University of Tubingen. He is a distinguished specialist of Immanuel Kant, of Hans-Georg Gadamer and of Martin Heidegger. His main areas of research include hermeneutics, phenomenology, classical German philosophy, the history of metaphysics and the philosophy of religion. We had the good fortune of having a discussion with him.

PENSÉES: What is philosophy? What led you to work in this domain?

JEAN GRONDIN: I would like to start by thanking you for this interview, especially since dialogue is the very essence of philosophy, whether it be the dialogue we have with others, or the incessant dialogue of the mind with itself. What is philosophy? I have absolutely nothing against its classic definition: the love of wisdom. Do you know of a better one? What led me to it? The fact that I do not have wisdom, and therefore, I wish to learn from those that do, or that have more than I. We come to philosophy because we are passionate about the questions it raises. It is certainly not the job opportunities that attract us to it. So there is a risk for those who go into philosophy. Man is a question, an enigma for himself, and we find interlocutors, as fascinating as they are rigorous, who help us pose this question and, sometimes, even outline some answers.

PENSÉES: Man is a question! That is beautifully phrased. Dialogue, question and answer are indeed at the very heart of the philosophical endeavour, and are – as some have seen – always already operational

as the true starting-point of philosophy. Any question, any answer, is situated in a dialogue that pre-exists them. This fundamental element is an important aspect of hermeneutical thought, particularly that of Hans-Georg Gadamer, which you know well. Can you tell us a few words about your mentor and about the hermeneutical movement, which has particularly interested you?

GRONDIN: Beautifully phrased indeed, so it cannot be from me. Augustine phrased it that way in the tenth volume of his *Confessions*. Gadamer has indeed been another great mentor to me. It is difficult, however, to only say "a few words" [about Gadamer], and the same can be said about the hermeneutic movement. What to say and where to begin? Gadamer is first and foremost a tremendous philosopher, whom I have had the opportunity to meet and get to know over the years. He is a penetrating defender of humanism, the meaning of which he reminds us at the beginning of his chef-d'oeuvre *Truth and Method*; the book's main idea being that our conception of truth, and of knowledge, is much too determined by the rather narrow perspective of methodical science and its technical goals, and without our realizing this completely. It is undeniable that truth is very often understood through this prism, which of course has uncontestable legitimacy, because it embodies a privileged route to reach the truth. However, Gadamer claims that it is not the only one, that there is a truth in practical wisdom (or in prudence), in acquired experience, in art, in history, and in philosophy itself. He is right in saying that the privilege of methodical knowledge tends to make us forget these experiences of truth, which *Truth and Method* helps us rediscover. Gadamer does this by drawing inspiration from Greek, Medieval and Modern philosophy, and the entire hermeneutical tradition, ancient and modern, which he brought to general attention. Hermeneutics, in short, is the discipline that reflects upon the great principles of interpretation, most notably in disciplines which always have to do with interpretation, such as exegesis, law, history and literature. But it was a little-known discipline before Gadamer. One of his great

merits is to have made it known and to have made it an unavoidable philosophy.

PENSÉES: This putting into question of the very notion of truth, as absolutely given and objective, seems to be at the heart of your questioning. Any 'truth' is, in fact, traversed by the fabric of interpretation, and this even in the so-called 'exact' sciences, within which the notion of interpretation has become more and more central, at least since the beginning of the 20th century. There are no 'naked' facts, so to speak: such is one of the lessons of hermeneutics. To what do you attribute this increased attention to the hermeneutical character of philosophical thinking, always subject to interpretation?

GRONDIN: Never, never would I say that it's a matter of putting the very notion of putting the notion of truth into question, quite the contrary! Where did you get that? I was speaking more of a rediscovery. Truth remains thought as adequation of the intellect and the real, and to me; there are indeed facts, whether naked or dressed. But you are right, sometimes it indeed does take a little bit of interpretation to get to them.

PENSÉES: The questioning of the notion of truth I suggested does not mean that the notion itself is rejected, quite the contrary: however, by bringing to light its conditions of possibility, it seems that we must wrap our heads around the less naïve idea that truth is always in part dependent on a given interpretative framework, and of a method that opens access to it. It is in this sense that I mean hermeneutics puts the notion of truth into question. The path to truth is thus always stems from a game of interpretation; it wells up from a process of questioning which must itself be explained and not simply taken for granted. If indeed there are facts, and if truth is always thought as *adæquatio intellectus et rei*, it must still be said that after the hermeneutical turn, the notion of interpretation becomes crucial to the very possibility of true discourse. I wanted to get to the constant interaction between truth, on the one hand, and method, on the other.

GRONDIN: In what sense is all truth dependent on a frame of interpretation? And must we really speak of a frame? Who exactly knows this frame? Is not the best interpretation the one that does not let itself be limited by any frame and that simply opens to what is to be understood? You see, now I'm the one asking questions. I suppose such things are allowed in philosophy. Now, concerning the tie between truth and method, you're right to speak of an interaction: it is often truth itself which commands the manner by which to approach it. Aristotle had already seen this. In this way, hermeneutics opposes itself to the Cartesian idea of a universal method that would be the condition of all truth. Every field shall have its own method, every object having its own requirements, to which thought and, by that very fact, its methodical approach, must conform.

PENSÉES: Indeed, if there is any discipline where addressee can become the addressor, it is within philosophy. It is this dialogical character of philosophy that provides philosophy with all its dialectical power.

GRONDIN: And to our discussion its meaning.

PENSÉES: As you have said, following Gadamer, it would seem that our interpretative frames have had the effect of restricting of conception of truth. Moreover, it is indeed this pragmatic and technocentric conception of knowledge that seems dominant, nowadays. In this, are we the inheritors of the modern Cartesian project? What's more, do you think that we could go so far as to say that the hermeneutical project overturns Descartes' foundationalist project and its (perhaps excessive) emphasis on the methodical acquisition of knowledge?

GRONDIN: In fact, Descartes is not the main target of Gadamer's critique, at least not directly. The conception of method espoused by Descartes – namely, the deductive method according to which knowledge must stem from a first certainty – is very different from the one which has come to dominate modern science; the method of which rests less, or not at all, on a primitive certitude, rather than on

Descartes' idea that the results of knowledge ought to be observer-independent and founded on observation. This works for the natural sciences, but can we simply apply this model to the social sciences, where the researcher's own implication in his knowledge of his field is non-negligible? It is this productive contribution of the researcher towards what he understands that Gadamer attempts to underscore, with the conception of understanding as a fusion of horizons (of the interpreter and the thing interpreted). You are right to say that it is Descartes, who without a doubt, has anchored this priority of method into our minds. However, everything is not just a question of method, even in Descartes, maybe even especially in his thought. The requirements of method proceed quite clearly (if you don't mind me saying…) from an doubting and hesitating *ego*, in a quest for orientation, meaning and truth (recall that *cogitare* comes from the classical Latin term for 'to hesitate', that is to say, to 'co-agitate' possibilities). Methodical truth thus presupposes the implication of the thinker, and more fundamentally still, finitude. One has to be think oneself quite clever, moreover, to believe that one can simply criticize Descartes with impunity without taking into account his entire corpus. It would be like criticizing Dante or Moliere.

PENSÉES: Mr. Grondin, I would sincerely like to thank you to have agreed to this interview. Our exchange was most interesting. Your answers beautifully illustrate the way in which the thinker himself is involved in his questioning. Do you have a few closing words you would like to address to the philosophers of tomorrow?

GRONDIN: Do not forget those of yesterday. They opened the way for you, and allowed you to think. Be wary of complacency, of school-line disputes, and of sectarian chapels, all strangers to a durable philosophy. Do not be too disappointed if the beautiful discipline of philosophy is not recognized in what is called the 'real world' (where two immense philosophical notions slumber). As Descartes said when he spoke of the tree of philosophy: essential philosophy, or metaphysics, questions the roots of what is, and these roots, it goes

without saying, are not always visible. Do not be too discouraged if the echoes of your work (the genuine echoes, of course) seem modest: philosophical ideas take time to make their way. All things considered, what is important is not immediate success, but the thought of tomorrow, which is in your hands. Thank you for taking the initiative to have this conversation.

PENSÉES: Thank you, Mr. Grondin, for this fusion of horizons.

Derrida and Merleau-Ponty on Husserl: The Heterogeneity in Tactile Auto-Affection

EMMA RYMAN
McGill University

INTRODUCTION

Merleau-Ponty, as a phenomenologist, is an inheritor of the tradition of metaphysics and works, in a sense, within it. As such, there are great disparities between his work and the work of Derrida, who frequently aims his deconstruction at the phenomenological project itself. The two philosophers are, in fact, typically considered to be "highly different, and even paradigmatically opposed" (Reynolds 2004, p. xv). However I will argue in this paper that there is a deep affinity between Derrida and Merleau-Ponty regarding the non-purity of tactile auto-affection. Both Derrida and Merleau-Ponty deal with the impossibility of pure auto-affection in response to the views regarding bodily sensation put forth by Husserl in *Ideas II*. Husserl uses the example of tactile auto-affection, as the immediate manifestation of the body proper to itself as such, to argue for the privileged position of touch. Derrida, in *On Touching – Jean-Luc Nancy*, responds to Husserl and argues against the possibility of pure tactile auto-affection. I will argue that, despite his critiques of Merleau-Ponty in *On Touching – Jean-Luc Nancy*, Derrida's views on the non-privileged position of touch and on the impurity of auto-affection bring his thought very closely in line with that of the later Merleau-Ponty. Merleau-Ponty, in *The Visible and the Invisible*, argues that touch and vision are fundamentally intertwined, and thus can be read as rejecting a hierarchy of the senses. Furthermore, he argues that auto-affection always involves hetero-affection due to the hiatus between touching and being touched. For these two reasons, I argue that there is indeed an interesting continuity between the respective philosophies of Derrida and Merleau-Ponty.

HUSSERL ON THE BODY AND AUTO-AFFECTION

In *Ideas II*, Husserl explores the role of the *Leib* (the body proper/ the Body) as the perceptual organ of the experiencing subject. The *Leib*, understood as the body *qua* the animated flesh of a human or animal, is to be distinguished from the *Körper*, understood as the body *qua* physical material.[*] Husserl explains that bodily sensations are *doubled* in auto-affection, such as when one hand touches the other. This is because when two body parts come into contact with each other, "each is then precisely for the other an external thing that is touching and acting upon it, and each is at the same time Body" (Husserl 1989, p. 153). The Body is thus constituted both as a physical thing with certain properties like extension, colour, and so on, and also as a sensing thing, which perceives sensations like warmth, texture, and so on. These sensations are experienced as *localized* in the Body, distinguished by the location of their appearance on the Body and also belonging phenomenally to the Body. Husserl (1989, p. 154) refers to these localized sensations as effect-properties of the Body, arising where and when the Body is touched.

In the case of touching-oneself, the sensations produced are more complicated than those produced when one touches other objects in the world. When one touches an external object, a double sensation is produced since the Body both experiences the tactile qualities of the external object and also experiences itself as something that is touching the external object. In self-touching, this double sensation is apprehendable in a double way. Husserl explains, "[I]n the case in which a part of the Body becomes equally an external Object of an other part, we have the double sensation (each part has its own sensations) and the double apprehension as a feature of the one or the other Bodily part as a physical object" (Husserl 1989, p. 155). This analysis of self-touching as producing a double apprehension of body-as-object and body-as-sensing-thing illustrates, according

---

[*] For a more detailed explanation of this distinction, see Husserl (1989, p. xiv-xv).

to Husserl, a difference between touch and sight. The eyes, Husserl explains, are not capable of such pure auto-affection. This is because the eye does not see itself during sight, and one eye can also neither touch itself nor the other eye to produce the double sensation experienced in touching-oneself.[*]   Simply put, "I do not see myself, my Body, the way I touch myself. What I call the seen Body is not something seeing which is seen, the way my Body as touched Body is something touching which is touched" (Husserl 1989, p. 155).

That is not to say that the eye is not a field of localization of sensations. Rather, it is a field for the localization only of touch-sensations instead of seeing-sensations – the eye can be touched immediately by the hand, but it cannot see itself immediately. In this sense, Husserl grants the auto-affection of touching-oneself a privileged position over seeing-oneself. Husserl (1989, p. 156) applies similar reasoning to hearing-sensations, which he claims involve the ear, but are not in fact localized to it. He concludes from this cursory examination of different types of sensations that,

> "*[L]ocalization of sensings* is in fact something *in principle different from the extension of all material determinations of a thing*. The sensings do indeed spread out in space, cover, in their way, spatial surfaces… [But the] sensing which spreads over the surface of the hand and extends into it is not a real quality of a thing" (Husserl 1989, p. 157).

The touching-sensation of the hand is the *hand itself*, not simply a state of the hand *qua* material thing. The hand is more than just a material thing by virtue of its ability to touch, and this is because it is through touch that my hand can manifest itself immediately to me as part of my Body. Again, Husserl (1989, p. 158) uses this immediacy

---

[*] Husserl (1989, p. 155n1) also rules out the possibility of seeing oneself in a mirror as pure auto-affection because looking in a mirror involves indirect judging.

of self-manifestation to illustrate a "privilege of the localization of the touch sensations… [because each] thing that we see is touchable, and, as such points to an immediate relation to the Body, though it does not do so in virtue of its visibility." Husserl (1989, p. 158) argues that a subject who only possessed the sense of vision "*could not at all have an appearing Body*" because it is tactility, and everything that is localized with the sensations of touch, that originarily constitutes the Body. The Body becomes something more than just a material thing only by incorporating tactile sensations via the localizations of sensations *as* sensations. This localization in the body is the precondition for the existence of all sensations and appearances. Thus, Husserl establishes the primacy of touching over the other sensations, both on the basis of its role as the sense that originarily constitutes the Body *qua* Body, and also on its intuitive immediacy, as exemplified in the auto-affection of self-touching.

DERRIDA'S CRITIQUE OF HUSSERL

In 'Tangent II' of *On Touching – Jean-Luc Nancy*, Derrida (2005) provides an analysis of Husserl's treatment of touching and auto-affection in *Ideas II*. Derrida (2005, p. 162) summarizes Husserl's views on the body, noting that the assertion that the distinction between *Körper* and *Leib* based on the condition of touch is why "the phenomenologist credits touch with an absolute, unparalleled, and grounding preeminence." However, Derrida questions Husserl's distinction between seeing and touching, which Husserl uses to illustrate the primacy of touching. As we have seen, Husserl believes that the "difference between the two "senses" lies in the self-relation of touch (and therefore its reflexive phenomenological evidence), which is *immediate*, *spontaneous*, *direct*, *intuitive*, and without equivalent in a mirror or mediation" (Derrida 2005, pp. 170-171). When I touch myself, according to Husserl, there is no mediation and no insinuation of alterity. This is not true for the eye, which cannot auto-affect without mediation. Therefore, as previously discussed, it lacks the

potential to experience a "fully intuitive, direct, and synchronous" double sensation (Derrida 2005, p. 171).

Derrida attempts to undermine this privileging of touch over sight on the basis of auto-affection by calling into question the purity of touching-oneself, and consequently, the possibility of immediate experience of the body *qua* Body (that is, the body *qua* the body proper). Derrida (2005, p. 179) does not wish to deny the possibility of the experience of tactile auto-affection, but rather determine whether or not this experience is "*haunted*, but *constitutively* haunted, by some hetero-affection related to spacing and then to visible spatiality". Derrida argues that there is a certain exteriority required for the touching-touched experience. The duplicity of the double sensation of touching-touched necessitates that the outside of what is touched announce itself as "foreign to the 'touching' and the 'touched' sides of the phenomenological impression at the same time" (Derrida 2005, p. 175). Without this experience of exteriority, there would only be one thing experienced – only the experience of being touched or only the experience of touching – and there would thus be no double apprehension. Therefore, Derrida (2005, p. 175), concludes, "[I]t is necessary that the space of the material thing – like a difference, like the heterogeneity of a spacing – slip between the touching and the touched, since the two neither must nor can coincide if indeed there is to be a double apprehension." This is not to say that, in the experience of auto-affection, I am not both the thing touching and the thing touched. Rather, some not-I – be it space, extension, and so on – insinuates itself between my experience as toucher and as touched. If this were not the case, then I would be unable to posit myself as "I" and distinguish myself from any other material thing; I would be indistinct from all other extended material. As such, there is a sense in which hetero-affection is the condition of possibility for auto-affection, rendering the experience of touching-oneself as a type of "auto-hetero-affection" (Derrida 2005, p. 180).

Derrida goes on to argue that a condition of this extension in space is that what is touching and what is being touched – in this case, my hand and my other hand – is visible, or potentially visible. This visibility, or this exposure to the outside, which is another way of saying *exteriority*, undermines the possibility of reducing touching-oneself to a pure experience of the Body as such. Consequently, Derrida (2005, p. 180) raises the question,

> "When Husserl seems to draw a line between, on the one hand, pure auto-affection of the body proper in the "double apprehension" of the touching-touched, and, on the other hand, the hetero-affection of sight or the eye… shouldn't [he] rather distinguish between several types of auto-hetero-affection without any pure, properly pure, immediate, intuitive, living, and psychical auto-affection at all?"

If there is always an element of hetero-affection in the auto-affection of touching-oneself, and furthermore, if the condition of tactility, which is spatial extension, is also the condition of visibility, then there is no reason to posit the purity of auto-affection nor to privilege touch over vision. Nor, Derrida (2005, pp. 179-180) continues, is there a reason to accord a privilege or priority to any sense. Thus, Derrida shows that tactile auto-affection contains within it a necessary detour through visibility and hetero-affection, which undermines the very privilege and purity that Husserl attempted to establish in his analysis of touching-oneself.[*]

---

[*] The continuity, or perhaps disparity, between Derrida's treatment of pure auto-affection in *On Touching – Jean-Luc Nancy* and in *Voice and Phenomenon* is a subject that is beyond the narrow scope of this paper. For discussion on this topic, see 'Chapter 2: With My Hand over My Heart, Looking you Right in the Eyes, I Promise Myself to You… Reflections on Derrida's Interpretation of Husserl' from Lawlor's 2006 book *The Implications of Immanence: Toward a New Concept of Life* and Durie's 2008 article, "At the Same Time: Continuities in Derrida's readings of Husserl."

DERRIDA AND THE LATER MERLEAU-PONTY

Merleau-Ponty also examines the experience of auto-affection throughout his work, arguing too against the privileging of touch and against the purity of tactile auto-affection. In a text that he worked on up until his death, *The Visible and the Invisible*, Merleau-Ponty (1968, p. 133) presents his final views on perception. He claims that perception is made possible when,

> "[M]y hand, while it is felt from within, is also accessible from without, itself tangible, for my other hand, for example, if it takes its place among the things it touches, is in a sense one of them, opens finally upon a tangible being of which it is also a part. Through this crisscrossing within it of the touching and the tangible, its own movements incorporate themselves into the universe they interrogate…"

Thus, Merleau-Ponty follows Husserl in using auto-affection, and specifically tactile auto-affection, as the basis for understanding perception in general. Yet, Merleau-Ponty goes on to say that this experience of tactile auto-affection is no different for vision except for the fact that the information that vision gathers does not belong to it, but to another sense (that is, touch). However, this distinction between senses, according to Merleau-Ponty, is crude. Every experience of the visible partakes in the field of the tangible – a movement of my hand creates just as much difference in what I see as it does in what I touch. The visible spectacle belongs to the touch neither more nor less than do the "tactile qualities" (Merleau-Ponty 1968, p. 134). As such,

> "We must habituate ourselves to think that every visible is cut out in the tangible, every tactile being in some manner promised to visibility, and that there is encroachment, infringement, not only between the touched and the touching, but also between the tangible, and the vis-

ible, which is encrusted in it, as, conversely, the tangible is not a nothingness of visibility, is not without visual existence" (Merleau-Ponty 1968, p. 134).

Thus, in *The Visible and the Invisible*, Merleau-Ponty departs from Husserl's privileging of touch over vision, and instead emphasizes the intertwining of the tangible with the visible. This non-privileging of touch seems to be reflected in Derrida's aforementioned arguments surrounding extension. Recall that Derrida (2005, pp. 174-175) claims,

"And it is there, precisely because of extensio, because of visibility and the possibility at least for the hand to be seen, even if it is not seen (a possibility involved in the phenomenological content of the sensible impression), that manual touching – even just touching my other hand – cannot be reduced to a pure experience of the purely proper body."

Derrida's critique of the purity of tactile auto-affection thus seems to follow a logic that is similar to that of Merleau-Ponty. Both philosophers argue against the legitimacy of reducing the experience of touching-oneself to a pure auto-affection of the Body, albeit from different theoretical bases. It is thus somewhat surprising that Derrida is so critical of Merleau-Ponty.[*] Throughout 'Tangent III' of *On Touching – Jean-Luc Nancy*. Derrida attempts to undermine both Merleau-Ponty's interpretation of Husserl's *Ideas II* and also his views on the relationship between the senses.[†] Derrida (2005, p.

---

[*] Much of Derrida's criticism of Merleau-Ponty is in regards to interpretation of Husserl's arguments on the means of (in)direct access to the bodies of others – an issue that is beyond the scope of this paper, although it is another interesting point of contrast and comparison between Husserl, Derrida, and Merleau-Ponty. See, for example, Husserl (1989, pp. 165-180); Merleau-Ponty (1964, pp. 165-176) and Merleau-Ponty (1968, pp. 140-6); Derrida (2005, pp. 187-197).

[†] For instance, Derrida focuses heavily on what he deems to be Merleau-Ponty's (1964) problematic misinterpretations, or perhaps violent reinterpretations, of Husserl's views

206) states that although Merleau-Ponty gives a great weight to the intertwining of the senses, he "never excludes a hierarchical order from it, and then confers on vision a heavy primacy." He argues that Merleau-Ponty actually privileges vision, allowing it to pave the way for touch, which nonetheless conditions it. This is a perplexing lesson to take from Merleau-Ponty's writing on the senses. Indeed, the only real support that Derrida gives for this assertion is a small part of a passage from *The Visible and the Invisible*. Derrida (2005, p. 187) quotes Merleau-Ponty as claiming, "[S]eeing is being premeditated . . . the visible body provides for the hollow whence a vision will come, inaugurates the long maturation at whose term suddenly it will see, that is, will be visible for itself." The full text of this passage in fact states,

> "In spite of all our substantialist ideas, the seer is being premeditated in counterpoint in the embryonic development; through a labour upon itself the visible body provides for the hollow whence a vision will come, inaugurates the long maturation at whose term suddenly it will see, that is, will be visible for itself, will institute the interminable gravitation, the indefatigable metamorphosis of seeing and the visible whose principle is posed and which gets underway with the first vision. What we are calling *flesh*, this interiorly worked-over mass, has no name in any philosophy" (Merleau-Ponty 1968, p. 147; my emphasis).

I quote the entirety of this passage in order to show the bizarre in-

---

on auto-affection in 'The Philosopher and His Shadow' in *Signs*. In this text, Merleau-Ponty states that in tactile auto-affection, "The relationship is reversed, the touched hand becomes the touching hand, and I am obliged to say that the sense of touch here is *diffused* into the body…" (Merleau-Ponty 1964, pp. 166-7, quoted in Derrida, 2005, p. 187). Derrida objects to this interpretation of the sense of touch as *diffused* into the body in auto-affection. On the contrary, Husserl's notion of touch-sensation is one that is *localized* to a point in the Body, not diffused. According to Derrida (2005, p. 190-205), this is just one example among many of Merleau-Ponty's questionable interpretive gestures regarding Husserl's views on sensation.

terpretation to which Derrida has subjected it. This passage, which Derrida takes to illustrate a privileging of vision over touch in *The Visible and the Invisible*, is clearly referring to *flesh* – a new concept that Merleau-Ponty (1968, p. 140) uses to denote something that is "not matter, [but rather] is the coiling over of the visible upon the seeing body, of the tangible upon the touching body, which is attested in particular when the body sees itself, touches itself seeing…" The concept of flesh includes within it both visibility and tangibility. The *visible* body that is meant to provide for the hollow whence a *vision* will come is the same *tangible* body that provides for the hollow whence a *touch* will come. It is thus unclear why Derrida draws from this passage evidence that Merleau-Ponty endorses a hierarchy among the senses that privileges vision. Of course, from the very title of the book, it is clear that *The Visible and the Invisible* emphasizes vision and the visible more than it does touch and the tangible. Nonetheless, an acknowledgment that Merleau-Ponty focuses on vision is not in itself a positive argument for Merleau-Ponty's alleged privileging of vision over touch in a hierarchy of the senses.

The only other support Derrida (2005, p. 206) offers for this interpretation of Merleau-Ponty is a passage from one of his earliest works, *The Phenomenology of Perception*, where Merleau-Ponty (1962, p. 234n1) claims,

> "[T[he senses should not be put on the same basis, as if they were all equally capable of objectivity and accessible to intentionality. Experience does not present them to us as equivalent: it seems to me that *visual experience* is truer than *tactile experience*, that it garners within itself its own truth and adds to it, because its richer structure offers me modalities of being unsuspected by touch."

While this is, admittedly, a clear statement on Merleau-Ponty's part of a privileging of vision, it seems obvious, based on the passages previously explored from *The Visible and the Invisible*, that the later

Merleau-Ponty had overcome this unnuanced, hierarchical view of the senses in his later work. Although he still focuses his attention on vision, the later Merleau-Ponty constantly endeavours to make clear the ineluctable intertwining of touch and vision. Thus, since Derrida has failed to provide any compelling evidence that the later Merleau-Ponty subscribes to a hierarchy of the senses, the force of Derrida's arguments surrounding the privileging of vision in *The Visible and the Invisible* is defused.[*] In fact, there is a sense in which Derrida too recognizes the possibility of non-privileging within *The Visible and the Invisible*. Derrida (2005, p. 205) acknowledges that it is "as rich a text as it is heterogeneous, and all the more profuse since it stays on the move and undecided with respect to all these alternatives [between seeing and touching, the invisible and the untouchable, and so on]." Not only can we read Merleau-Ponty as not subscribing to a hierarchy of the senses in *The Visible and the Invisible*, but we can also understand Derrida's views on the non-priority of any given sense as being in accord with the later Merleau-Ponty's philosophy, or at the very least, profoundly connected.

A further point of similarity arises between Merleau-Ponty and Derrida regarding their views on "auto-affection" as "auto-hetero-affection".[†] Merleau-Ponty (1968, p. 147) explains that in auto-affection,

> "My left hand is always on the verge of touching my right hand touching the things, but I never reach coincidence;

---

[*] Derrida (2005, p. 199) also refers to Dastur's claim that Merleau-Ponty confers an "exorbitant privilege" to vision, but he does not present a positive argument in support of this claim.

[†] This affinity between Merleau-Ponty and Derrida on the issue of auto-affection is mentioned in passing by Lawlor (2009, p. 220), where he states, "When Merleau-Ponty in *The Visible and the Invisible* describes the touching-touched relation, he discovers within the relation of auto-affection a kind of distance, or even a blind spot, which makes auto-affection heterogeneous. The claim that auto-affection is always hetero-affection brings Merleau-Ponty's thought very close to that of Derrida."

the coincidence eclipses at the moment of realization, and one of two things always occurs: either my right hand really passes over to the rank of the touched, but then its hold on the world is interrupted; or it retains its hold on the world, but then I do not really touch it – my right hand touching, I palpate with my left hand only its outer covering."

In tactile auto-affection, the experience I have of my hand as touching never completely coincides with my experience of my hand as touched. There is a hiatus between my hand touched and my hand touching, just as there is, for example, a hiatus between my voice heard and my voice uttered. Were there no hiatus, that is if my experience of auto-affection was indeed immediate, then I would fuse with what is experienced – my two hands fusing into one experienced thing – making the experience itself impossible. As Merleau-Ponty (1968, p. 122) puts it, "The moment my perception is to become pure perception, thing, Being, it is extinguished." It is the necessity of this hiatus, this distance, this non-coincidence, or what Lawlor (2009, p. 226) refers to as a *blind spot*, which makes auto-affection heterogeneous, and makes it fundamentally so.

CONCLUSION

Derrida's views regarding the heterogeneity of auto-affection can be understood as consonant with those of the later Merleau-Ponty. As we have seen, Derrida critiques Husserl's notion that auto-affection is somehow pure and unmediated by arguing that a difference, like the heterogeneity of a spacing, must insinuate itself between the touching and the touched for the double apprehension in auto-affection to take place at all. The upshot of Derrida's argument is virtually identical to that of Merleau-Ponty's. Both insist on the absolute necessity of an element of alterity or heterogeneity within the experience of auto-affection for it to be experienced as such. Thus, despite

proceeding from different philosophical perspectives, both Merleau-Ponty and Derrida undermine the purity of tactile auto-affection in a similar manner.[*]

Although the philosophies of Merleau-Ponty and Derrida are far from identical, there are nonetheless two interesting points of continuity between their respective philosophies with regard to their treatment of auto-affection in reference to Husserl's *Ideas II*. First, after arguing that Derrida's criticism of Merleau-Ponty's privileging of vision is unsubstantiated, I showed that it is plausible to read *The Visible and the Invisible* as emphasizing the continuity of the senses rather than the primacy of vision. As such, Derrida's rejection of a hierarchy of the senses can be seen as in line with Merleau-Ponty's thought. Second, I argued that Derrida's critique of Husserl, which appeals to the heterogeneity necessary for auto-affection, is very similar to the hiatus that Merleau-Ponty identifies within the touching-touched experience. Thus, despite their differences, there are striking continuities between Merleau-Ponty and Derrida with regard to the status of auto-affection in perception.

---

[*] Although Derrida does not seem to explicitly acknowledge this strong affinity, he does at one point gesture towards the importance of Merleau-Ponty's thought with regards to auto-affection. He credits Merleau-Ponty's thought for opening up a pathway for the thinking of Merleau-Ponty's time, and of his own time. Nonetheless, Derrida (2005, p. 211) remains critical of Merleau-Ponty because he never carried out a "more powerful reformalization of his discourse in order to thematize and think the law under which he was thus placing himself – always, in fact, and all things considered, preferring "coincidence" (of coincidence with noncoincidence) to "noncoincidence" (of coincidence with noncoincidence)." Whether or not Derrida is correct in this assertion is, unfortunately, beyond the scope of this paper.

REFERENCES

Derrida, J. 2011. *Voice and Phenomenon*. Translated from French
     by L. Lawlor. Illinois: Northwestern University Press.

Derrida, J. 2005. *On Touching – Jean-Luc Nancy*. Translated from
     French by C. Irizarry. Stanford: Stanford University Press.

Durie, R. 2008. At the Same Time: Continuities in Derrida's
     readings of Husserl. *Continental Philosophy Review,* 41(1),
     pp. 73-88.

Husserl, E. 1989. *Ideas Pertaining to a Pure Phenomenology and to
     a Phenomenological   Philosophy: Second Book*. Translated
     from German by R. Rojcewicz and A. Schuwer. Dordrecht:
     Kluwer Academic Publishers.

Lawlor, L. 2006. *The Implications of Immanence: Toward a New
     Conception of Life*. New York: Fordham University Press.

Lawlor, L. 2009. Becoming and Auto-Affection (Part 2): Who Are
     We? *Graduate Faculty Philosophy Journal*, 30(2), pp. 219-
     237.

Merleau-Ponty, M. 1962. *The Phenomenology of Perception*.
     Translated from French by C. Smith. London: Routledge
     & Kegan Paul Ltd.

Merleau-Ponty, M. 1964. *Signs*. Translated from French by R. C.
     McCleary. Evanston: Northwestern University Press.

Merleau-Ponty, M. 1968. *The Visible and the Invisible*. Translated
     from French by A. Lingis. Illinois: Northwestern University
     Press.

Reynolds, J. 2004. *Merleau-Ponty and Derrida: Intertwining
     Embodiment and Alterity*. Athens: Ohio University Press.

Inventing Logic:
The Löwenheim-Skolem Theorem and First- and Second-Order Logic

VALÉRIE LYNN THIERREN
Université de Montréal

INTRODUCTION

On its own, the Löwenheim-Skolem theorem – as well as the ensuing Skolem Paradox – is of scant mathematical, logical, and even philosophical significance. It is of no consequence to practicing mathematicians, admits of extant resolutions at the logical level and is sufficiently vague to support all sorts of philosophical arguments.[*] Nevertheless, much enquiry and ink has flowed on the matter, and we intend here to contribute to this muddy river, all the while trying to avoid waxing poetic or waning technical. In this essay, we purport to use the paradox as a paradigm by which we mean to evaluate the role of first- and second-order predicate logic in a post-foundational context. As there is no reason to prefer one kind of logic over another unless we specify to what purpose we intend to use it, we shall argue that if we want mathematical logic to axiomatize, describe and couch the language of informal mathematical practice, then second-order logic yields more *intuitively appropriate* models. As well, since the Skolem paradox is a problem purely for standard model-theoretic semantics, our secondary purpose will simply be to show through this example how modern model-theoretic results can illuminate our current understanding of what logic is, as well as of what we conceive the role of logic to be in the greater scheme of human understanding. Indeed it is our over-arching intention to ponder the inherent limitations and, paradoxically, the intrinsic openness of human thought.

---

[*] Most notably, within Skolem's own expoundings on the subject. For a succinct rundown of this topic, we refer the reader to George (1985).

1.0 WHAT LOGIC?

Philosophic logic is generally understood as the formalization of vernacular languages. There are in fact many types of such formalizations: syllogistic, propositional and modal logics all attempt to capture the syntactic structure of speech patterns -- that is, of thought itself.[*] Much of modern logic, however, arose in a mathematical context, more precisely in the context of the foundational crisis of the early twentieth century. As a preamble to our discussion on the (epistemo)logical implications of the Löwenheim-Skolem theorem on the axiomatization of mathematics, we shall first briefly review the nature of mathematical logic as a peculiar branch of logic. As well, we shall find it worthwhile to touch upon the distinction and relation between model-theoretic semantics and proof-theoretic syntax which is crucial to our argument. Finally, we shall review in this section the characteristics of first- and second-order logic (and by extension, higher-order logic) – which is ultimately the object of our paper.

1.1 POST-FOUNDATIONAL MATHEMATICAL LOGIC

Nowadays, mathematical logic can be defined as the attempt to achieve an adequate formalization of mathematical language; alternatively, it can be considered as the study of the deductive and expressive power of formal theories. Historically, mathematical logic owes its inception to the search for the foundations of mathematics. The crisis in the foundations of mathematics was wrought with vigorous debate about purely abstract entities such as the uncountably infinite sets evoked by Cantor's Theorem. During the foundational crisis, logicism aimed to reduce all of mathematics to fundamental logical 'laws' of thought, to be expressed through an ideal and closed

---

[*] We must here note that there is a strong positive feedback effect between the thought expressed through language and the language that underlies the thought.

formal logical system with explicit axioms sufficient to characterize all abstract mathematics -- initially considered to be part and parcel with set theory. Set theory, however, was rife with contradictions and paradoxes; the first axiomatic wave in the field of logic focused quite exclusively on smoothing out and formalizing Cantor's paradise – work which was taken up notably by Hilbert, Russell, Whitehead, Zermelo, Skolem, and Fraenkel (Crossley et al. 1972, p. 5; Ferroiros 2001, p. 471).

Despite the obsolescence of foundational studies proper, this work remains a driving force behind much research in mathematical logic and many logicians and mathematicians still explicitly or implicitly harbour hope that some kind of non-absolute foundation can still be achieved. Most notably, the ghost of Hilbert's programmes can still be glimpsed in debates into the 'nature' of mathematical logic. As it had once been hoped that mathematics could be reduced to logic, it seems that the ideal that logic should be devoid of any mathematical content or presuppositions is still circulating: logic should formalize mathematics, but if mathematics impregnates logic, we arrive at a vicious circle. On the other hand, it is not clear that there is a non-ideological reason why mathematics must rest on a foundation that is itself non-mathematical in nature.[*] Hence, while we shall not discuss foundational studies in this paper, we do note that many prejudices still seem to hold: 'logic' logic is equated with first-order logic, and 'mathematical' logic is equated with second-order logic – which

---

[*] In fact, the reverse statement sounds rather absurd barring convincing non-ideological reasons. Evidently, much has been made of the presence or lack thereof of a border between logic and mathematics; generally, however, mathematical logic is considered a sub-discipline of both mathematics and logic. While some maintain that even a fuzzy border is a border – a border that restricts each discipline to its own sphere – we are not of that opinion. If anything, the haziness of this border renders it *a priori* 'undecidable' to which domain a 'foundation' (or, even a simple axiomatization) of mathematics must be sought. Regardless, as Boolos has remarked, if all the other sciences presuppose themselves, why shouldn't mathematics? See Boolos (1975, p. 517), Gauthier (1976, pp. 294-6), and Shapiro (1999, pp. 51-54).

is itself conflated with set-theory; Many important philosophical issues concerning the nature of logic itself surround the debate between first- and second-order logic – to which we shall soon turn our attention, after having first delved into the semantics/syntax dichotomy that pervades modern logic (Shapiro 1985, p. 742).

1.2 MODEL THEORETIC SEMANTICS

Although modern logic arose during the *Grundlagenkrise der Mathematik*, its crux occurred after its falter. We thus find it impossible to delineate the scope of what logic is, what it can do, and what we would like it to do, without wording it in proof- and model-theoretic terms. All mathematical logic consists of two parts: a formal language governed by recursive proof-theoretic syntax and an informal language governed by descriptive model-theoretic semantics. The formal language is ideally uncontaminated by the content or meaning of the linguistic sentences it codifies: the archetypal formal system is a sturdy logical skeleton that wavers not in the winds of philosophical opinions. The informal semantics is the interpretation of the theory that reflects unto the language its truth conditions.

As the Löwenheim-Skolem theorem is a model-theoretic problem, it is this aspect of the logical conundrum that retains our attention. Models accomplish the linguistic reference-fixing of the consequence relations delineated by the deductive system. At times, these models may reveal the defects in our logical systems that are in need of adjustments, though at others they may expose flaws in our broader epistemological systems instead. Indeed, since Tarski, model-theory has often been understood as a theory *about* truth. This is perhaps why model-theory has come to dominate and define mathematical logic.[*] But if "formal language is to model-theory what language is

---

[*] Harold Hodes even goes so far as to claim "truth in the model is a model of truth" (Shapiro 1999, p. 44).

to the world" (Shapiro 1999, p. 44), and if the world 'contains' the truth that the language seeks to reconstruct, then an analysis of the model brings to light not only what we know, but also what we *can* know, what we *cannot* know, as well as *how* we know (Gauthier 1976, pp. 121-2 and 218; Shapiro 1999, pp. 43-4, 56).

1.3 FIRST- AND SECOND-ORDER LOGIC

In this paper, we are concerned solely with first- and second-order predicate calculus with standard model-theoretic semantics. First-order predicate logic (henceforth, 'FOL') consists of a given non-empty domain *d* within which a countable infinity of quantified variables range over the individual elements. Standard first-order model-theoretic semantics are fundamentally characterized by completeness, compactness and the Löwenheim-Skolem theorem (henceforth, 'LST'). After the metalogical conclusions of Gödel's enquiries, FOL became highly valued for its ability to generate a full deductive system. This appraisal reflects the strong proof-theoretic tradition that suffuses all of mathematics: there can be no results without proof. For these reasons, first-order calculus is still the *de facto* fundamental logical language (Ferreiros 2001, p. 470; Gauthier 1976, pp. 64, 123; Manzano 1996, p. 112; Tharp 1975, pp. 4, 7).

Second-order predicate logic (abbreviated 'SOL') with standard model-theoretic semantics have interpretations wherein additional 'second-order' quantified variables range over all of the subsets of elements of a given non-empty domain *d*, along with the standard 'first-order' variables that range over its objects alone. Most early logicians, like Zermelo, initially worked purely with second-order logic. First-order logic was an offshoot of the original predicate logic and was later championed by logicians like Skolem for mostly ideological reasons – but practical reasons as well: first-order logic is more easily and crisply manipulated (Crossley et al. 1972, p. 5; Ferreiros 2001, p. 471). Since Gödel's incompleteness theorem shows

that compactness and LST inherently fail in second-order axiom systems, no standard model-theory of SOL can produce a complete deductive system (Gauthier 1976, p. 64; Shapiro 1985, p. 714; Shapiro 1999, p. 42).[*]

Although it is generally considered that FOL trades expressive power in the name of securing clear epistemic gains, there is a damper: a complete deductive system alone neither provides knowledge nor generates new knowledge (Gauthier 1976, p. 218; Jané 1993, p. 67; Shapiro 1999, pp. 44-5). Nonetheless, the other side of the coin is that SOL is overly expressive: it has been claimed that, by saying everything, it in fact says nothing. Indeed, though its incompleteness lends itself to dizzying heights of expressive potential, its deductive mechanisms are cumbersome and often bewildering. For this reason, it is SOL that must bear the burden of proof (Gauthier 1976, p. 123; Jané 1993, p. 71; Manzano 1996, pp. 5, 60-2, 112).[†]

2.0 WHAT PARADOX?

One of the earliest meta-mathematical results to arise from model-theoretic research into FOL was the Löwenheim-Skolem theorem. In this section, we shall first explore the Löwenheim-Skolem theorem, including the upward and downward expansions of the theorem which we owe to Tarski. Then, we shall touch upon the ensuing Skolem paradox that arises in the semantic interpretations of first-order models,[‡] as well as its basic philosophic import. The paradox

---

[*] On the other hand, some non-standard second-order models *do* have all three of these properties – most notably, Henkin semantics. However, all the same objections (including the ones that derive from the presence of LST) that we will address later hold for these second-order models (Gauthier 1976, p. 123; Shapiro 1985, p. 715).

[†] The problem with FOL is that its strong syntax cannot generate results that constitute new knowledge, whilst the problem with SOL is that its strong semantics do not allow logicians to show which of its disparate results constitute knowledge.

[‡] Given that LST does not hold in second-order theories, Skolem's paradox cannot arise in second-order model-theoretic semantics.

not being a true antinomy, and it being but of minor significance to mathematicians, we shall describe a few solutions of which it admits within the framework of metamathematical logic. The impact of LST on our understanding of first- and second-order predicate logic, as well as on logic and epistemological possibilities themselves, will be addressed in the next chapter.

2.1 THE LÖWENHEIM-SKOLEM THEOREM

As we have seen, the Löwenheim-Skolem property – along with completeness and compactness – is a hallmark of first-order models. Quite simply, the Löwenheim-Skolem theorem states that any theory that is consistent (i.e., that has a model) has a countable model. Whilst prior to Gödel's completeness and compactness theorems,[*] the Löwenheim-Skolem theorem is nevertheless its immediate consequence, as it states that if a formula is satisfiable (i.e., has a model), then it is satisfiable within a countable domain. However, given Gödel's accompanying compactness theorem, LST follows from the completeness theorem in such a way that it is possible to derive the upwards and downwards Löwenheim-Skolem-Tarski theorems (or 'LSTT'). The original theorem, as refined by Tarski, can thus be formulated in two versions: a) the upward Lowenheim-Skolem-Tarski theorem and b) the downward Löwenheim-Skolem-Tarski theorem. The upwards LST states that if a theory has any model of infinite size, then that theory also has a model whose domain is the same size as an infinite set *A*; in other words, a satisfiable set of sentences always has a model of a greater infinite cardinal. As a corollary, the downward LST proves that if M is a model of cardinality K and if λ is a cardinality smaller than K, than M has a submodel of cardinality

---

[*] Gödel's completeness theorem states that FOL mechanically produces all the valid logical formulas that follow from its axioms, such that any logical expression is either satisfiable or refutable in a model. His compactness theorem states that a set of sentences in FOL has a model if and only if all of its finite subsets also have a model (Crossley 1972, p. 7).

λ which satisfies the same theory as M itself; in other words, every satisfiable sentence has at most a countable model (Shapiro 1985, p. 714; Schoenfield 2001, p. 79).

By Lindström's theorem – one of the pioneering technical results of model-theory – LST is required if we want a strong first-order theory.[*] However, the presence of the Löwenheim-Skolem property means that first-order theories cannot manage the cardinalities of its infinite models, in such a way that its models are not categorical – that is, they are not isomorphic. As a consequence, while LST plays an important role in proving the strength and completeness of FOL, it nevertheless has the serious unintended consequence of paving the way for Gödel's 1931 incompleteness theorem wherein (through a process now known as the 'arithmetization' of logical syntax) Gödel proved that FOL was not strong enough to either prove or disprove the formulas of classic mathematics, such as arithmetic. FOL is incomplete in regards to the theory of natural numbers based on Peano's well established axioms,[†] and this is a highly problem-

---

[*] Lindström's theorem establishes that if a model of any given logic *L* has either Löwenheim-Skolem-Tarski property and is also compact (and, therefore, complete), then *L* is equivalent to first-order logic. FOL is then equated with the maximal (strongest) logic possible – but only given these provisions. Indeed, the characterization of the 'strength' of FOL as defined by this theorem is dependant upon the presumption that LST (and compactness, and completeness) are *essential* characteristics of sound logic: if they are not, then invoking Lindström's theorem amounts to begging the question (Tharp 1975, pp. 4-9). The question of elucidating whether or not LST is a necessary feature of logic (which would then imply that FOL is the strongest logic period) is by no means one we are prepared to answer, but it is a question we will confront as we weigh the undesirable consequences of the Löwenheim-Skolem property against those of a logical system that does not possess it.

[†] The incompleteness theorem also had considerable ramifications on the burgeoning discipline of proof-theory – which has led, amongst other things, to the universally accepted (yet unprovable) Church-Turing thesis which states that the absolute undecidability of a formula can be decided via algorithms. This theorem was crisply trailed by a whole slew of undecidability results concluding that the whole of mathematics as well as a host of basic problems – such as the Continuum Hypothesis itself (when combining Gödel's and Cohen's consistency results) – were essentially undecidable. With these results arose for the first time a fundamental problem of consistency within elementary number theory. Of

atic situation. Indeed, LST has been deemed the first of the modern incompleteness theorems, casting shadows on the assumption that FOL can be the kind of strong logic that mathematical theory and practice can rely on (Kleene 1952, p. 427).

A final note: because of the lack of completeness and compactness of second-order theories, some theoreticians will admit only FOL as 'valid' logic. However, the lack of a Löwenheim-Skolem property in second-order models of these theories is not actually considered a problem: rather, it is thought of as a favourable characteristic – for reasons we shall see shortly. As such, while the Löwenheim-Skolem theorem and Skolem's paradox cannot quite be considered a logical deal-breaker, it is a tipping-point – a genuine model-theoretic problem. Of course, to understand how model-theoretic semantics can influence our understanding of the formal theories themselves and further delineate the scope of logical enquiries, it shall be necessary to describe the ensuing Skolem paradox.

2.2 SKOLEM'S PARADOX

The Löwenheim-Skolem theorem leads to a simple paradox: the appearance of a seeming contradiction between the Löwenheim-Skolem theorem (which proves that if a formula or list of formulas is satisfiable, then it is $\aleph_0$-satisfiable) and Cantor's theorem (which proves the existence of non-denumerable sets of cardinality $2^{\aleph_0}$). Overtly, since we have a countable theory that proves the undenumerability of some sets, how can the ensuing model account for the existence of uncountable objects within its countable domain? For we then have countable models for axiom systems that are intended to structure uncountable domains – which is unseemly. The

---

course, all of these considerations are also of consequence within the mathematical and philosophical discourse on the nature of truth (Boolos 1975, p. 523; Crossley 1972, pp. 7-10; Kleene 1952, pp. 300-1, 317-8, 436). As seminal as Gödel's incompleteness results are, we regret that we can provide no more space to the subject in this present paper.

paradox confronts us with the reality that any axiomatization of set theory employing a denumerable amount of formal axioms will fail to render and characterize the absolute concepts of set, power-set, bijection, non-denumerability, etc., which are the very fundamental set-theoretic notions that we intend mathematical logic to explain (Kleene 1952, p. 427; Kleene 1971, pp. 326-9).

Furthermore, as already alluded to, the upward and downward LST reveal the basic non-categoricity of first-order models, further reinforcing the non-absoluteness of set-theoretical notions in FOL. This has led to a (relatively) marginal meta-mathematical and philosophical conclusion called Skolemism. Skolemism is the idea that set-theoretical notions are inherently relativistic: what is non-denumerable in one interpretation of a formal system may be denumerable in another, as there is no prior absolute definition of non-denumerability; instead, uncountability is a property relative to a given model, not of the formal system.[*] For the Skolemite, the paradox proves the model-relative nature of all cardinality results. Skolem's paradox is thus a problem specifically for model-theory: it paradoxically reveals that a complete and compact formal axiom system produces many interpretations – including unintended ones that rub against the grain of the formal theory we thought we were building up. LST thus inherently provides the impetus for non-standard models (Crossley et al. 1972, pp. 6, 29; Ferreiros 2001, p. 472; Kleene 1952, p. 427).[†] Of course, to a certain extent, this is natural: the scientific method has a sneaky way of crushing our hunches, hypotheses and our desires;

---

[*] Though, to be fair, it is hotly contested whether we even have a prior solid reason to believe that these are notions that can be rendered at all, given set theory itself is in need of a formal 'foundation' (Jané 1993, pp. 78-83; Shapiro 1999, p. 58). However, it is not our aim to engage in foundational debates, and so this will remain here a moot point.

[†] The first of these models was constructed by Skolem himself. It is through the work of Henkin however that such non-standard models are revealed to be an inherent consequence of the model-theoretic relativity of complete and compact FOL with Löwenheim-Skolem properties (Crossley et al. 1972, p. 29).

but applied to model-theory, the consequences of LST seem to reveal that this is a problem *within* even the most rigid application of the scientific method itself.

2.3 SOLUTIONS TO SKOLEM'S PARADOX

Skolem's paradox has definite solutions available, whether we like these particular resolutions or not. As it is not an antinomy (but is, rather, a mere incongruity), it may be brushed aside by informal mathematic practice and, indeed, by much of formal mathematics. And therein lies a distinction between the logician and the mathematician, for while the mathematician is content to work with logical concepts and models for purely investigative and constructive reasons, it is the logician who must concern himself with the ontic starting point of the concepts and models invoked (Klenk 1976, pp. 476, 479). However, logic itself cannot produce any ontic knowledge without subscribing to an underlying ideological position postulating its prior existence. Therefore, we shall explore here only the solutions pertinent to the model-theoretic semantics of first- and second-order theories.

2.3.1 WITHIN FIRST ORDER LOGIC

The upward Löwenheim-Skolem-Tarski theorem fix

The aforementioned upward and downward versions of LST comprise what may be considered a reflection schema, wherein the upward version correlates to the ascending reflection of all acceptable models of higher infinite cardinality and the downward theorem represents a descending reflection of all admissible models of lower (in)finite cardinality. Viewed in this light, a dialectical interpretation through the upward and downward variants saves the Lowenheim-Skolem theorem from itself, with no need to step outside a first-order theory. Thus, within FOL, the upwards Löwenheim-Skolem-

Tarski theorem may be invoked to dispel the paradox and show that the continuum can be built up through an ascending reflection of the countable models which underlie it. Indeed, this may well account for the hypothesis that denumerable models are sufficient to describe set-theory (Gauthier 1976, pp. 297-8; Klenk 1976, pp. 475, 479, 485). However, the upward LSTT has its own host of drawbacks, which shall be succinctly addressed in the next section.

The misinterpretation interpretation

This solution can also be dubbed the 'much ado about nothing solution'. Indeed it may be claimed that while the model-theoretic interpretation understands the existential and universal quantifiers to range only over the domain of M, the observing logician 'intuitively' understands instead that '$\exists x$' and '$\forall x$' to range over the *entire* set-theoretic universe. In other words, while only $\aleph_0$ elements can be 'observed' within the model's domain, the full spectrum $2^{\aleph_0}$ can be 'observed' from without. However, within the countably infinite model, the infinite sets actually are countable – they may really be placed in bijection with the natural numbers. The apparent paradox arises simply because this bijection does not actually occur from within the perspective of the model. The denumerable model thus internally satisfies the notion of 'non-denumerable set', though the model is in fact countable to the external observer of the model. This is how a countable model can be said to sufficiently describe the continuum. Indeed, in retrospect, it seems rather trite to say that if M is countable – and the paradox is thus rendered somewhat banal. Of course, this is still an unexpected and unsettling explanation to the logician as it leaves him or her with no compass by which he may explain his own notion of (non-)denumerability (Crossley et al. 1972, pp. 6, 29; Kleene 1952, p. 426).

The non-standard model-theoretic semantics way out

Of course, it can also simply be accepted as fact that there are standard and non-standard models. It could very well be that we

commonly use a particular arithmetic, but that there might be a manifold of consistent arithmetic that we could also work with in a sound manner. In fact, mathematicians do frequently build and explore non-standard models of theories, some mathematicians become specialists of these models, and some believe that one or more non-standard models are in some shape or form 'better' than non-standard ones.[*] Skolem's paradox isn't an antinomy, it is then just an unexpected consequence: one we must learn from. In actuality, there is no Platonist or intuitive conception of 'set-theory' as either an ideal fact or a fixed intuitive notion to which we ought to fit our model-theoretic semantics. Rather, we develop and construct what set-theory really is or is not based on the results achieved through axiomatization, model-theory and, especially, proof-theory. Nothing *beyond* what is proved is of any scientific substance; if a consistent, complete and compact logic necessarily entails LST and the Skolem 'paradox', then *that* is logic, *that* is reality. There is simply no other way to proceed than to attempt to solve 'known' problems – by any means necessary.[†] The costs of this resolution will be elucidated in the fourth section.

2.3.2 WITHIN SECOND-ORDER LOGIC

---

[*] 'Better' need not mean absolutely better. A non-standard model might simply be better at capturing a certain use or a certain possible application of a theory. As an analogy, we can truthfully say that the duodecimal notation is 'better' for counting, that a non-simple continued fraction is 'better' at expressing π, or that the binary system is 'better' for building Turing machines. Again, 'better' always refers to a certain context.

[†] The fix just mentioned smoothly recoups the constructivist position. In fact, according to McCarty and Tennant, Skolem's paradox does not actually arise at all in constructivist practice. Indeed, it is a basic underlying assumption among constructivists and some intuitionists that Cantor's theorem and the notion of 'non-denumerability' have no place in logic, as they cannot be constructed in the rigorous sense of the word. At best, it is possible to suspend belief as to the matter, but it is still widely believed within constructivist circles that even countable infinity is *de juris* unconstructible. Evidently, such a position entails that LST – and therefore Skolem's paradox itself – is unconstructible (McCarty & Tennant 1987).

As we know, LST only applies to first-order formal theories. It does not apply either to second-order structures, or to the informal mathematics that first-order model-theoretic semantics intends to capture. LST does not apply to SOL because SOL does not itself produce LST, and thus it does not produce the infamous paradox. Furthermore, second-order models of ZFC were shown in 1930 by Zermelo (himself an early champion of SOL) to accurately interpret the informal set-theoretic notions of cardinality and power-set, something with which first-order axiomatizations struggle. Furthermore, SOL characterizes up to isomorphism the classic theory of natural numbers as established by Peano's axioms (Shapiro, Second-Order Lang.). The details of this (and other benefits) incurred by SOL due to its lack of LST shall be dealt with later on in the next section.

3.0 WHAT SEMANTICS?

As we have seen, LST and Skolem's paradox are intrinsically tied to FOL. But as we have also seen, FOL is intrinsically tied to LST by Lindström's theorem.[*] Though the paradox is not a strict contradiction of the sort that poses a strong problem for mathematics itself, it does highlight some of the inescapable limitations of any first-order axiomatization of mathematics. Most immediately, the question arises as to whether there is an inherent contradiction in first-order model theory, and what this contradiction tells the respective logical and mathematical communities. Indeed, it seems to us that the paradox gives us a clearer idea of what logic can or cannot do, but most importantly, whether the interest lies in a secure but restrictive FOL, or in a treacherous but expressive SOL. This section shall thus review what the Löwenheim-Skolem theorem can tell us about first- and second-order logic.

---

[*] See footnote on p. 34.

3.1 THE LÖWENHEIM-SKOLEM THEOREM IN FIRST-ORDER LOGIC

A limitation imposed on FOL by the upward and downward LST is that it is in fact impossible from a model-theoretic perspective to say what the cardinality of the underlying model is: though we can say that it is infinite, we cannot say in an absolute way whether the model is actually imbedded in an uncountable model. Of course, the upwards Löwenheim-Skolem-Tarski states that if the string of formulas has a model, it admits also of models of every infinite cardinality K – but this does not allow us to deduce whether the countable model is characterizing an uncountable domain which is itself 'really' uncountable (Crossley et al. 1972, p. 29; Shapiro 1985, pp. 716-9).[*]

Given the ubiquity of the concept of non-denumerable sets in mathematics, it is expected of mathematical logic that its language allows for an expression of this property, as well as many others that are essential to mathematical theory. Yet applied to first-order logic, LST is tantamount to stating that FOL is not powerful enough to generate an enumerating function capable of expressing the concept of non-denumerability: as such, what appears to be non-denumerable within M, is in fact denumerable from without (Kleene 1952, p. 426; Klenk 1976, p. 475). In fact, the presence of LST renders FOL's semantics inadequate even to allow for a satisfactory expression of the basic arithmetical principle: Peano's axiom of induction. Also, proof-theory being inexorably intertwined with model-theoretic semantics, LST thus has indelible consequences on the strength of FOL as a deductive system: despite completeness, the range of what it can prove is so limited as to render it rather an absurd choice for the description of mathematical practice (Boolos 1975, p. 521; Jané 1993, pp. 68, 70; Shapiro 1985, pp. 716, 722, 727).

---

[*] This idea can be expressed something like stating that there exists an uncountable 'universe' outside the domain of the countable model at hand, a universe of which the model captures only a piece (Crossley et al. 1972, p. 70). But then that raises the question: can we provide a model for this universe?

3.2 THE LÖWENHEIM-SKOLEM THEOREM IN SECOND-ORDER LOGIC

Because LST fails in SOL, its model-theoretic semantics allow for the convenient and sound expression of countable and uncountable sets, since non-denumerability *is* characterizable in second-order model-theoretic semantics: the formula expressing the uncountable domain simply admits of no countable model. As such, SOL interprets 'correctly' our intuitive set-theoretic ideas. Furthermore, the crucial axiom of induction in ZFC, as well as Peano's axioms and the axioms of separation and replacements, are also best expressed at the second-order.[*] Arithmetic, real and complex number analysis, Euclidean spaces, real vector spaces... the list of theories uniquely and exactly expressible through second-order model-theoretic semantics is far too extensive to be dismissed out of hand (Boolos 1975, pp. 521-2; Jané 1993, pp. 68-71, 79-81; Manzano 1996, pp. 4-5; Shapiro 1985, pp. 722, 727, 729-30).

Another benefit conferred unto SOL by its lack of LST is its categoricity results. Because SOL is not compact and is not imbued with the Löwenheim-Skolem property, its models are categorical; for example, it can characterize the sets of natural and real numbers – and many other infinite structures – up to isomorphism (Shapiro 1985, p. 714; Shapiro 1999, p. 43). This may very well prove to be a greater strength than the syntax of FOL; after all, to correctly describe a structure is to describe it categorically, and second-order model-theoretic semantics does this for denumerable and non-denumerable structures, as well as most of the structures of classical mathematic theory. Categoricity also seems more essential than completeness when considering the interdisciplinarity of the mathematical subfields, as exemplified by the common practice of embedding specific

---

[*] Of course, this and other axioms may be 'replaced' with first-order axiom schemes. However, this solution requires that we postulate an infinity of axioms (to the unique second-order axioms), a move which is quite preposterous in light of a formal system that cannot even express what it is that is meant by 'infinite' within its own model- and proof-theory (Boolos 522; Jané 79-80; Shapiro, Second-Order Lang. 725-6)!

theories within another, most commonly set-theory – for which second-order axiomatization has thus far proven indispensable (Boolos 1975, pp. 523-5; Jané 1993, p. 70; Shapiro 1985, pp. 716-7, 722, 728, 739). After all, even if a theory were to be explainable through FOL, if its models are not categorical, not much is gained.

So then what is the advantage of limiting the scope of logic solely to FOL when it is well known that its expressive power is inherently poor and inadequate to express even the basic givens of classical mathematics? Well, SOL's semantic power are themselves impugned: it has been claimed that all the answers it claims to hold are merely hypothetical for lack of a complete deductive system to draw them out (Jané 1993, pp. 81-4; Klenk 1976). As both FOL and SOL have their advantages and disadvantages, it is impossible to say which one is the 'better', unless we specify what we are using the logic for. The answer seems deceptively simple, but is in reality quite treacherous: we want mathematical logic to reproduce our intended models. However, since the Löwenheim-Skolem property prohibits the isomorphic categorization of our intended infinite models, FOL inherently leads us to consider unintended models – and this is a serious problem for most logicians.

4.0 WHAT INTENDED MODELS?

Amongst other things, Skolem's paradox highlights a visceral meta-mathematical impasse: what constitutes an intuitively acceptable model? After all, we are wont to think that mathematicians have some grasp of the theory they are in the midst of elaborating; however, a naive Platonist vision of the mathematical objects that populate the reality 'behind the model' is anathema to the modern, level-headed theoretician. But are intended models 'real' models? Are unintended models aberrations of truth? Must we reject a formally true theory if it yields unintended models? Are intended models philosophically anachronistic? None of these questions can be acceptably answered

lest we first ponder the question: what is model theory (and mathematical logic itself) a theory of?

4.1 MODEL-THEORY AND INFORMAL MATHEMATICS

We here mean to deepen our previous discussion: why should mathematical logic prefer semantic expressiveness over completeness? What exactly are we trying to express? And why are not proof-theory and formal axiom systems enough to express it? Of course, mathematical logic is engaged in an inexhaustible dialogue with standard informal mathematical practice.[*] Intended models are just the models that are intended by practicing mathematicians; it is their informal theories and models that mathematical logic intends to formalize. For while it may turn out that the informal theories are found wanting in some aspect or another, it is to informal mathematics that we must turn to if we are to give any meaning to the term 'intended model'. We want the models of our formal systems to coincide with the intended objects and relations we mean to describe – ideally categorically. These *are* the original models, the ones we hold all others up to (Shapiro 1999, pp. 45, 48).

After all, informal mathematics is both the beginning and end point of mathematical logic: it not only provides the immediate impetus and *raison-d'être* of logic, but it is also hoped that logic has applications – namely, that it is useful and illuminating to proof-seeking practitioners. It is through informal mathematics that we can conceive, however vaguely, of what we mean by the standard intended model that unites the majority of mathematicians and logicians (Jané 1993, pp. 68-9; Shapiro 1985, pp. 725-6). As such, what this practice says and to a certain extent how it does it constitutes our standard. We shall now contrast the correlations between first- and second-order models with standard semantics with the informal mathemati-

---

[*] Perhaps sometimes aptly referred to as pre-formal mathematics.

cal practice that grounds it.

4.2 THE LÖWENHEIM-SKOLEM THEOREM AND INFORMAL MATHEMATICS

As noted, LST is not generally considered to be a problem for practicing mathematicians. But despite the fact that LST admits of resolutions both at the first- and second-order level, it shall nevertheless prove instructive to enquire more into the preferences of the mathematicians themselves. As has been stated, FOL is generally considered incapable of axiomatizing some of the most elementary branches of mathematics – including set-theory, which is crucial to nearly all branches of the discipline. Not surprisingly, what FOL cannot axiomatize is what it can't model: infinite structures. This is where LST can be considered the tipping-point, especially if one considers the fact that the language of infinity is crucial to our modern understanding of mathematics and physics (Shapiro 1985, pp. 714-5, 719, 739). LST is an issue for model-theoretic semantics, and where FOL fails the test is precisely in the semantic section. The semantics involved in informal mathematical practice -- even such well-understood notions as finitude, mathematical induction, minimal closure and well-founded relations like the predecessor and 'less than' relations (which may all be easily constructed with second-order formulas) – simply cannot be expressed in first-order axiomatizations. FOL neither resembles our mathematical structures nor seems useful to the working mathematician looking for guidance in far-off places like philosophy, logic and meta-mathematics (Shapiro 1985, pp. 722-4, 727).

Of course, there are reasons to prefer FOL, but in a post-foundational epoch one must re-evaluate the traditional authority of first-order theories, especially if one considers that foundational studies are ultimately about founding pre-formal mathematics. While SOL can no more provide a foundation for all of mathematics, it does shed more

light on mathematical theory and practice than its first-order counterpart. Second-order model-theoretical semantics simply provide better models of the infinite structures that are vital to modern mathematics. Also, they capture better the meanderings and sinews of the semantics of ordinary mathematic discourse;[*] indeed, it seems that informal mathematics, which dispenses totally of the LST, functions semantically at the second-order level of logic. Indeed, the universal preference for second-order languages by practicing mathematicians can be interpreted as their universal rejection of Skolemism (Jané 1993, p. 67; Shapiro 1985, pp. 720, 727, 739; Shapiro 1999, pp. 44, 62).

5.0 INVENTING LOGIC?

As we have seen, intended models refer to the *use* of particular terms and sentences in informal theory, but that such use is subject to some kind of evolution (especially when reflected to itself by a formal theory's model) is quite banally evident and inevitable. While the model then means to capture the use of informal notions, the model itself still must be semantically interpreted to gauge whether this use is in fact adequately expressed. Here, there seems to be a curious interplay: while we want our formal models to capture something of our intended models, the formal model itself does not ever point to the intended model and, in fact, provides no real means by which we may gauge the veracity of our intended models, or even what we ourselves mean by our vague idea of the intended model. Furthermore, without semantic interpretation, it is difficult to assess what we are building formal theories of, and perhaps most importantly, why we are building them in the first place. The intent of this section is to offer up a few further musings on what model-theoretic semantics can tell us about our broad meta-mathematical, logical and epistemological aims.

---

[*] Indeed, of ordinary vernacular discourse itself (Klenk 1976, p. 483).

5.1 MODEL-THEORETIC SEMANTICS AND INTENDED MODELS

The Löwenheim-Skolem theorem confronts us with the facile way in which we rely upon unspoken agreements as to our intended models. Indeed, the relativity results inherent to LST can lend credence to the idea that logic might just be a purely formal science. Indeed, given the LSTT, Skolem's paradox proves *neither* the existence nor the non-existence of non-denumerable sets as the first-order model actually allows sentences to be interpreted either in the denumerable or the non-denumerable domain (Klenk 1976, pp. 476, 479). The interpreter's choice is then guided by other than purely formal considerations – such as one's prior commitment to the ontological existence or characteristics of non-denumerable sets. All that the LST actually states is that the models generated by FOL cannot characterize, cannot *say* anything about whatever the hypothetical structure of non-denumerable sets might be (Klenk 1976, pp. 479, 484). Under this light, the results of the application of the LST to FOL are not criteria by which one can measure whether the intuitively intended model of sets has been achieved – rather, it begs such questions as: what 'intended model'? Why not unintended models? Why not the model that we completely, if counter-intuitively, built?

While the answer to such questions lie outside the scope of our essay, we find it sufficient for the moment to postulate that if we mean to describe mathematics – but especially if we mean to 'found' mathematics – then our models had better do this; if this is the case, it simply will not do to have an axiom system incapable of reproducing the basic conceptual underpinnings of modern mathematics. Of course, our conceptual understanding might need adjusting but it will have to be adjusted in light of a theory within which it can discuss itself. Besides describing mathematical structures, it would also be nice if we could derive as well a model of mathematical activity itself. However, since FOL needs to replace simple second-order axioms (like the axiom of induction) with infinite axiom schemes, it seems that

even the syntax of FOL is insufficiently strong to deduce the most basic of the intuitions and principles that mathematicians work with. It thus seems that the informal logic of proof-theory itself is second-order – though we may with more or less success transcribe it to the first-order.

5.2 MODEL-THEORETIC SEMANTICS AND FORMAL SYSTEMS

Meta-mathematical speculation as to what *is* an intended model put aside, in actuality, formal logic *does* rely on informal mathematical theories and practice to guide its investigations: it is those theories that we want to generate in our models. However, modern axiomatics are not merely meant to codify and crystallize static states of knowledge or disembodied mathematical notions floating in the sky above. Or, rather, if this is what some intend it to do, then modern axiomatics also entails a host of unintended consequences. First and foremost, the formalization of informal theories often reveals unexpected epiphenomena – of which Skolem's paradox has struck us as a worthy paradigm. Because of such results, the formal theory itself morphs and adapts to these results through repeated intrusions and manipulations of its holdings. Sometimes, however, it is informal practice that is subtly or not so subtly modified by the observations and clarifications of model-theoretic results (Klenk 1976, pp. 480, 482).

As such, even though our mathematical hunches and techniques may need fine-tuning, a model in which classical mathematics cannot recognize itself will be very hard-pressed to provide answers, solutions and guidance to informal mathematics. The structural and formal approach is not sufficient to account for both the state and the needs of informal mathematics. Mathematics is a rigorous but inherently creative and ultimately intuitive discipline. Formalization is what happens *after* the edification of sufficient reasons warranting such a formalization, as well as sufficient guiding ideas. We cannot build a vacuum. In fact, we

are tempted to say that 'incompleteness' is part and parcel with the way humans effectively think – even about mathematics.

As such, logicians must take into account the symbiotic relation between informal practice and formal system if logic is to reflect mathematics, and if logic is to be a productive science – especially as proof-theory becomes more and more entangled with model-theory (Gauthier 1976, p. 293; Shapiro 1985, pp. 716-7; Shapiro 1999, pp. 46, 50, 56-7). The dichotomy between informal mathematics and formal logic has indeed become a tenuous premise to uphold (Klenk 1976, p. 482), one that not only sheds little light on meta-logical aims, but cuts off its supply. Strict logic has not reified itself, it has become more and more acutely aware of its limitations. If anything, modern logic has not become static, it has exploded into a convoluted host of standard and non-standard models, first-, second- and higher-order logics and set theories, etc.

5.3 MODEL-THEORETIC SEMANTICS AND EPISTEMOLOGY

We want logic sometimes for logic's sake but also for the sake of interdisciplinary research. We wish to apply logic in order to discover inalienable truths about the world. Logic *is* a branch of epistemology; it is a branch of philosophy that perhaps more than any other is the science of knowledge, or of measuring knowledge (Gauthier 1976, 296-7). While we do not state that either the Skolem paradox or any particular interpretation of a model has any ontic import,[*] we do state that we intend to capture at least some sort of ontological reality, whether or not we ourselves (can) capture this reality.[†] Although an ideal logic will be devoid of any content (i.e., will be

---

[*] No existence claims stem either from first- or second-order logic (Jané 1993, p. 68).

[†] The extant literature spawned by Putnam's "Models and Reality" (1980) attests to the desirability and general plausibility of such a thesis. However, Putnam's hypotheses are of no weight to our present discussion, which limits itself to the meta-mathematical implications of model-theory.

purely syntactical) (Boolos 1975, p. 517; Jané 1993, pp. 67, 72-4), this strikes us as a Platonic ideal of sorts for many reasons. We ought to apply logic to elucidate a domain, a structure, a world, a universe that is itself inherently populated – by us.

While logic does not reveal meaning, and even though it is a highly contested idea that anything *has* meaning, we nevertheless intend it to be a tool by which we can explore, purify and understand the meanings that we do ascribe to the constructed universes within and without ourselves. It is to this extent that we do not wish to reduce logical results to simple epistemological and ontological epiphenomena. If logic means nothing, why do we build it? For what would we want models, other than for gleaning, gauging and separating some sorts of truth? The model is a tool: it can tell us what does not work, what is contradictory, what is consistent, what is satisfactory, and what is merely interesting. It confronts with a certain externalization, a map of particular mental labyrinths through which we daily cut our paths. It is a creative tool for investigation, for exploration. It characterizes equally the true, the false, the hypothetical, the fantastical, the interesting. What we do with these representations, how we semantically interpret the given formal logic is how we link what could be but merely interesting observations with an even richer and more complex reality within and without ourselves.

CONCLUSION

In light of its inherent limitations, FOL is inapt to characterize and structure informal mathematical theory and practice. Seeing as how a vast portion of mathematical theories (as well as physical theories) require that they be imbedded in set theory, and seeing as how formal logic itself has a set-theoretic pedigree, the role of second- and higher-order logic is currently of greater import. This is not to say that the soundness of set-theory itself does not need to be further explored and expounded, but it is to say that a development and con-

solidation of SOL is to be recommended if one wants to found mathematics albeit in a non-absolute manner (*pace* Jané 1993, pp. 68, 74-5, 78, 85) and understand what mathematics and logic is about and what they can *say* about human understanding and cognitive functions. However much we might want and even need to formalize theories, it would be foolish and unproductive to ignore the inherent creativity involved both in mathematical and logical practice. And if what we want logic to do is formalize that reality as much as it is possible for it to do, SOL (much like non-standard models, quantum logic and multivalent propositional logic) is unavoidably logic (Gauthier 1976, pp. 294-5). It is certainly not all that logic is, but to disregard SOL and its substantial weight is to do a great disservice to logic and the development of a fundamental branch of knowledge in regards to the human experience.

REFERENCES

Boolos, G. 1975. On Second-Order Logic. *Journal of Philosophical Logic*, 16, pp. 509-27.
Crossley, J. 1972. *What Is Mathematical Logic?* London: Oxford University Press.
Ferreiros, J. 2001. The Road to Modern Logic – An Interpretation. *The Bulletin of Symbolic Logic*, 7 (4), pp. 441-84.
Gauthier, Y. 1976. *Fondements des Mathématiques: Introduction à une Philosophie Constructiviste*. Montreal: Presses de l'Université de Montréal.
George, A. 1985. Skolem and the Löwenheim-Skolem Theorem: A Case Study of the Philosophical Significance of Math ematical Results. *History and Philosophy of Logic*, 6, pp. 75-89.
Jané, I. 1993. A Critical Appraisal of Second-Order Logic. *History and Philosophy of Logic*, 14, pp. 67-86.
Kleene, S. 1952. *Introduction to Metamathematics*. New York: D. Van Nostrand Company.

Kleene, S. 1971. *Logique Mathématique*. Translated from French by Jean Largeault. Paris: Librairie Armand Colin.

Klenk, V. 1976. Intended Models and the Löwenheim-Skolem Theorem. *Journal of Philosophic Logic*, 5, pp. 475-89.

Manzano, M. 1996. *Extensions of First Order Logic*. Cambridge: Cambridge University Press.

McCarty, C. and Tennant, N. 1987. Skolem's Paradox and Constructivism. *Journal of Philosophic Logic*, 16, pp. 165-202.

Putnam, H. 1980. Models and Reality. *The Journal of Symbolic Logic*, 45, pp. 464-82.

Shapiro, S. 1985. Second-Order Languages and Mathematical Practice. *The Journal of Symbolic Logic*, 50, pp. 714-42.

Shapiro, S. 1999. Do Not Claim Too Much: Second-Order Logic and First-Order Logic. *Philosophia Mathematica*, 7, pp. 42-64.

Schoenfield, J. *Mathematical Logic*. 2001. Boca Raton: A K Peters/CRC Press.

Tharp, L. 1975. Which Logic is the Right Logic? *Synthese*, 31, pp. 1-21.

The Cognitive Equivalence Argument and Its Critics: A Critical Appraisal

LUKE KERSTEN
University of Ottawa

1.0 INTRODUCTION

The hypothesis of extended cognition (HEC) claims that cognition, at least in some cases, extends or spans across processes in the brain, body and environment at large. Proponents of this position claim support from a variety of different arguments.[*] However, there is one form argument which has had particular influence, what can be called the 'Cognitive Equivalence Argument' (CEA).[†] Arguments of this form hold something like the following: i) For any *x*, if *x* is equivalent to a cognitive process then *x* is itself a cognitive process ii) For some *x*, a process spanning, brain, body, environment, *x* is cognitively equivalent to a cognitive process iii) Therefore, *x* is a cognitive process.

The thrust of the argument is to claim that because there are no relevant differences between putatively cognitive processes on the one hand and processes spanning the brain, body and environment on the other, we should be motivated to call these further processes cognitive as well. In response to such arguments authors such as Frederick Adams and Kenneth Aizawa have invoked the need of a "mark of the cognitive" to distinguish cognitive from non-cognitive processes, and thus guard against cases of extended cognition.[‡] In

---

[*] Adams and Aizawa (2010) point out that proponents of HEC often make use of what they call "complementarity" and "evolutionary" arguments.

[†] Instances of the CEA can be found throughout the HEC literature (Clark 1998; Menary 2007; Noë 2004).

[‡] Adams and Aizawa outline two features that provide a mark of cognition. First, cognitive processes must involve non-derived representation. Second, they must be individuat-

other words, the mark of the cognitive is supposed to establish what can and cannot be legitimately called a cognitive process (Adams & Aizawa 2010, pp. 22-175).

Presently, I will attempt to analysis the mark of the cognitive as it is used both for and against the CEA. I argue for two claims. First, Adams and Aizawa's objections to Clark's 'Inga-Otto' and 'Futuristic Tetris' examples are question begging. Second, Clark's version of the CEA does not entail the conclusions of the HEC as Clark believes. In order to establish these claims, I offer a three-levelled analysis based on David Marr's levels of description. I aim to demonstrate that the true shortcoming of Clark's CEA examples and Adams and Aizawa's objections lie in their *incompleteness*. I will begin by outlining the form of Clark's CEA and then move to his examples.

2.0 CLARK'S VERSION OF THE CEA

Clark's CEA is a variation of the more general form I sketched above, it is what may be called a 'functional isomorphic argument' (Adams and Aizawa 2010, p. 150). It says: i) For all x, if x is functionally equivalent to a type of information processing then x is a cognitive process; ii) There exists an x such that x is a type of process spanning, brain, body, environment, that is functionally equivalent to a type of information processing; and iii) Therefore, x is a cognitive process. Or, more formally:

i) $\forall x\,(\mathrm{FE}(x) \to \mathrm{CP}(x))$
ii) $\exists x\,\mathrm{FE}(x)$
iii) $\exists x\,\mathrm{CP}(x)$

What is immediately noteworthy about Clark's version of the CEA is that it appeals to "functional equivalence". What is meant by this

---

ed by specific kinds of information processing mechanisms. These two marks will emerge in our discussion, as Adams and Aizawa attempt to rebuke Clark's CEA with them.

is that if information plays the same *functional role* in a process, in terms of guiding intelligent behaviour, then we can say there is cognitive equivalence (Clark 1998, p. 13).[*] This version of the CEA is distinct from the more general form previously outlined in that it stipulates a 'functional criterion' for cognitive equivalence. In short, if a process is functionally equivalent to a cognitive process then it is a cognitive process. We can note, however, that the functional criterion does not in itself need to be presupposed for the CEA to run. Nevertheless, the functional criterion is crucial here because it forms the standard for judging cases of cognitive equivalence according to Clark. With Clark's version of the CEA spelled out, I will now turn to Clark's first example of the CEA, the classic Inga-Otto example.

2.1 INGA AND OTTO

The Inga-Otto example asks us to imagine two individuals, Inga and Otto. Both of these individuals have the belief that they want to go to a museum, however, neither can remember where it is. Inga recalls the relevant information through brain bound biological memory, while Otto uses artefact memory, i.e., he stores information in a notebook which he checks when he wants to remember something. According to the orthodox view on cognition, Inga's case is a genuine cognitive process, while Otto's case is not. At this point Clark's CEA comes into play. It suggests that because both Inga and Otto's memory processing uses the information about the location of the museum in the same *functional role* to govern intelligent behaviour (i.e., get to the museum) we should be willing to count Otto's artefact memory as a cognitive process alongside Inga's biological memory (Clark 2010, pp. 228-9). In other words, insofar as there is an isomorphism between the information processing of Inga and Otto's

---

[*] Adams and Aizawa think Clark explicitly endorses something akin to this functional isomorphic argument when arguing for cognitive equivalence (Clark 1998, pp.8-14).

memory, as it governs intelligent behaviour, then we should count both cases equally as genuine cognition. To discount Otto's artefact memory in spite of this functional behaviour guiding equivalence is merely to demonstrate our biased brain bound view of cognition. Clark concludes, then, that if a process, such as Otto's artefact memory, spans the brain, body, and environment, and is a cognitive process, then we have a case of extended cognition.

Against Clark's first example Adams and Aizawa argue that there are *principled* differences between Inga and Otto such that cognitive equivalence does not obtain (2010, p. 136). Adam and Aizawa's object that Clark's appeal to functional equivalence, based on guiding intelligent behaviour, as a criterion for individuating cognitive processes, is misplaced. This is because we need to appeal to "underlying mechanisms" to individuate cognitive processes. To support this claim Adams and Aizawa attempt to outline a number of candidate mechanisms which could be used to individuate cognitive processes: mechanisms such as primacy and recency effect, memory recall, and depth of processing. They maintain that because these underlying mechanisms can be taken to differentiate Inga and Otto's memory then there can be no cognitive equivalence in the way Clark suggests. In other words, Adams and Aizawa seem to be objecting to Clark's characterization of the Inga-Otto case by referring to laws of cognitive psychology. They claim that these laws reveal principled differences between the processes at work in Otto and Inga when they recall the location of the museum. In particular, they claim to follow cognitive psychologists by saying that only those mechanisms which have been found to support cognitive processes should individuate processes. This is because, as it turns out, although these underlying mechanisms *could* have been similar, or even equivalent, as a matter of empirical fact the mechanisms are distinct. Thus, there cannot be cognitive equivalence. In short, the underlying mechanisms distinguish cognitive from non-cognitive processes and not merely the functional equivalence as Clark wants to suggest, therefore, the

CEA fails because the underlying mechanisms are not appropriately applied to individuate the memory processes between Inga and Otto.

This first objection seems deeply problematic. If Adams and Aizawa are appealing to the underlying mechanisms to differentiate cognitive processes then this seems to be question begging. If the underlying mechanisms are used as the criterion to individuate cognitive processes then it seems clear that only those processes supported by the underlying mechanisms will count as cognitive. However, claiming that Otto's memory is not cognitive because it is not supported by the proper mechanisms just assumes that cognitive processing – for example, memory – must resemble in its underlying mechanisms Inga's biological memory in the first place. In other words, if we use the underlying mechanisms, as Adams and Aizawa suggest, then Otto's artefact memory will clearly not count as cognitive, because it is not supported by the same mechanism as Inga's biological memory. In short, to suggest that Otto's artefact memory cannot *prima facie* count as cognitive is only to insist, but not argue, that Clark's functional criterion is the wrong one for individuating processes. However, Adams and Aizawa cannot simply point to the presence of mechanisms supporting the putatively cognitive cases, such as Inga's memory, and then say that these must therefore be the criterion for drawing a distinction between cognitive and non-cognitive processes, at least not without further argument.

Alternatively, perhaps Adams and Aizawa are not presupposing that the underlying mechanisms are the criterion for individuating processes. Rather, perhaps instead they are drawing attention to the fine-grained causal details between Inga's biological memory and Otto's artefact memory. Their claim then would be that Inga and Otto's memory processes are not equivalent because they are distinguished by the causal processes which realize them. However, if this is the objection then it again seems question begging. If Adams and Aizawa are attempting to show only that there are causal differences between the Inga's and Otto's memory processes, which in turn support the

underlying mechanisms, then this would be relevant only if these causal differences *prima facie* identified Inga's memory as cognitive and Otto's as not cognitive. Furthermore, Clark was not making the obvious claim that there are no principled *causal* differences. Rather, he was attempting to offer a model of explanation for gauging *which* principled differences should count as relevant to cognition; that is, functionally behaviour guiding information processing.

Adams and Aizawa seem to recognise this. They state, "Clark and Chalmers might charge us with looking at too fine a level of detail in comparing Otto and Inga, where we charge Clark and Chalmers with looking at too coarse a level of detail in comparing Otto and Inga" (2010, p. 140). However, after saying this Adams and Aizawa then go on to claim that we should appeal to the cognitive psychologists to take our lead in telling us "what to count as cognitively relevant differences" (2010, p. 140). However, Clark's task seemed not to be to question the empirical evidence produced by cognitive psychologists as to what underwrites cognitive processes causally in the brain, but rather to provide an explanation for what grounds to draw equivalence between putatively cognitive and other types of processes. It seems, then, that Adams and Aizawa's objections do not offer a relevant challenge to Clark's example, unless their mark of the cognitive has already been accepted. So, let us turn to Clark's second example of the CEA.

2.2 FUTURISTIC TETRIS

The second example of the CEA is Clark's futuristic Tetris game. Clark describes three possible cases of Tetris playing––the point of the game is to slot blocks into their correct fit. In the first case Clark maintains there is genuine cognitive problem solving because the case involves mentally rotating an image of a block to fit into the correct slot. In the second case the problem is solved by the externally activating a device, pushing a button, which then physically rotates

the blocks on a computer screen. Finally, in the third case, the task of slotting blocks into their correct fit is solved by a hybrid of the first and second cases. An external device, a neural implant, is activated which then mentally rotates the blocks for the agent. The point of the example is to show that if we would be willing to accept the first case as genuine cognition and plausibly by extension the third case, then we can further claim that the second case must also be cognitive. Or, in other words, Clark claims that because there are no 'principled differences' between (1) and (3), then by transitivity (2) follows. The information processing required to solve the problem is functionally equivalent in all three cases. If we accept this conclusion, then we seem to have, Clark suggests, an instance of extended cognition i.e. a process spanning the brain, body and environment (Clark 2010, p. 227).

In response to this second case of the CEA, Adams and Aizawa claim that there *are* principled differences between (1) and (2). These differences reside at the cognitive and causal levels. In (1) the agent deploys cognitive mechanisms to facilitate the task (i.e., mental representations). In (2) the task is not facilitated by such cognitive mechanism – that is, the blocks are not *representations* at all, rather, they just *are* blocks being rotated on a computer. Involved in (1) are cognitive processes supported by particular underlying mechanisms in the brain. These, however, are sharply distinct from the mechanisms involved in (2), where the blocks are rotated on the computer screen by electrons being fired at a phosphorescent screen. The point, then, is that the differences between (1) and (2) are all the attending cognitive processes involved in accomplishing the task for (1), and the distinct causal processes involved in (2). Hence, the Clark's 'no principled difference' claim does not hold because there are substantive differences between (1) and (2) such that there cannot be cognitive equivalence (Adams and Aizawa 2010, p. 142).

Similar to the previous response to the Inga-Otto example, Adams and Aizawa seem to be begging the question against Clark here. They

make a distinction between the cognitive and causal mechanisms involved in problem solving for (1) and (2). They maintain that in (1) problem solving is supported by cognitive mechanisms, e.g. mental representations, while in (2) problem solving is supported by causal mechanisms, e.g. electrons being fired at a phosphorescent screen. However, how can they appeal to underlying mechanisms like mental representation, which has already been assumed to be the mark of the cognitive, in order to distinguish cognitive from non-cognitive processes? It seems that was the point in question: what substantially individuated processes as cognitive. In other words, if the mark of cognitive is to play a role in distinguishing cognitive from non-cognitive processes, then it surely cannot be a role that *prima facie* decides empirically which cases are to count *as* cognitive – or else this would be making a stipulative definition of cognition and Adams and Aizawa clearly do not want to maintain that given their continuous appeals to following cognitive psychologists (2010, p. 144). Hence, again the objection is lacking a non-question begging flavour.

In summary, I have offered a refutation of Adams and Aizawa's objections based on what seems to be their question begging nature. What is noteworthy about this discussion is that it seems that Adams and Aizawa are putting forward their own definition of what makes a process cognitive, in opposition to Clark's CEA, and this is why their arguments against Clark have continued to fall short. The problem is, Adams and Aizawa cannot just point to their definition of what individuates cognitive from non-cognitive processes and then say that Clark's criterion is wrong: they must somehow show how their definition is superior to Clark's. And yet, they have failed to do this with any of their arguments. So, before we move to my analysis of Clark's CEA, let us first briefly look at how Clark responds to these objections. This will help clarify not only what Clark is claiming in his examples but also further explain why Adams and Aizawa fail to confute Clark.

3.0 CLARK'S RESPONSE

Clark responds to Adams and Aizawa's objections by questioning whether the mark of the cognitive really helps to distinguish what is "cognitive" in the way Adams and Aizawa supposed. He writes, "to argue from the other direction, and to identify cognitive processes as those that happen to characterize the neural activity of human agents, is to risk an unwarranted narrowing of focus (to the neural) and dangerous and unappealing chauvinism to boot" (Clark 2010, p. 93). He maintains that the only relevant mark of the cognitive is to draw attention to those processes which support "intelligent behaviour" (this is what the Otto-Inga or Tetris game examples highlight). This for Clark allows us to push what we should consider as cognitive from intracranial to transcranial processes, or extended cognition. He says for example, "what makes a process cognitive it seems to me, is that it supports intelligent behaviour"(Clark 2010, p. 92). Clark cannot see the force of Adams and Aizawa's objections because cognition, for him, is constituted by a complex coupling between processes intra and transcranially located, which are described by their functional poise guiding intelligent behaviour. Hence, the appeal to non-derived representations and underlying mechanisms to individuated cognitive processes is unconvincing for Clark.

I think there are two ways we can understand Clark's response. Either it is denial of the need for any mark of the cognitive or it is a putative rejection of Adams and Aizawa's mark of the cognitive in favour of his own. If we interpret these comments in the first way, then Clark's comments about intelligent behaviour guiding as a way to distinguish cognitive from non-cognitive process become unintelligible. This is because information processing guiding intelligent behaviour as a functional criterion seems to operate as distinguisher for the cognitive/non-cognitive divide – similar to Adams and Aizawa's appeal to underlying mechanisms. But, if Clark were rejecting the ability to distinguish cognitive from non-cognitive processes then it would make little sense to invoke intelligent behaviour production

by functional processing in order to response to Adams and Aizawa. I offer this interpretation because Clark repeatedly says he is not offering a "mark of the cognitive" (2010, p. 93). However, if this were the case then his version of the CEA would not work because it relies on a functional criterion to establish cognitive equivalence.

Alternately, if we take Clark as offering the second option then the functional equivalence of processes supporting intelligent behaviour becomes a "mark of the cognitive." Understood in this way, Clark's rejection rests on a rejection of Adams and Aizawa's criterion. This interpretation seems more in line with Clark's running of the isomorphic functional argument and why he seems so unconvinced by Adams and Aizawa's continuous invocation of their mark of the cognitive. In highlighting Clark's response, we can see how Clark and Adams & Aizawa talk across purpose: they both have distinct conceptions of the criteria for cognition.

The problem, then, is that there are two incompatible criteria on the table: Clark's functional criterion of guiding intelligent behaviour and Adam and Aizawa's causal/underlying mechanisms criterion. For this reason neither is compelled to take seriously the criterion employed by the other. This is why, although contrary to what Adams and Aizawa wish to maintain, Clark's instances of the CEA cannot be undermined by appeal to the underlying mechanisms, insofar as it already decides what does and does not count as a cognitive process. So how can we adjudicate this impasse? I think the answer lies in first characterizing both Clark's and Adams and Aizawa's criteria inside a larger frame of analysis.

4.0 LEVELS OF ANALYSIS

In order to answer why Adams and Aizawa fail in their objections and why the present examples of the CEA fail, we need to draw on a version of David Marr's distinction between levels of description. In short, Marr suggests that in order to understand complex informa-

tion processing systems we should adopt a three tiered theoretical perspective.

At top level we have cognitive functions, we can call this the computational level. At the computational level a system is explained in terms of what function it performs, i.e., what functions it computes, or its input/output behaviour. At the middle level we have algorithms or mathematical state transitions. The algorithmic level explains a system in terms of the computational procedures or algorithms by which it does what it does. And at the bottom level we have implementation. The physical or implementational level explains a system in terms of the physical mechanisms by which it operates (Horgan and Tienson 1993, p. 163). Taken together these levels of description constitute an explanatory approach to defining complex systems.*
The relevance of this explanatory model to our present concern is that this way of understanding mentality requires that each level be taken into account in order to explain cognitive phenomena. If we neglect one of the three levels of analysis then we run the risk of losing access to the theoretical tools at each level. In short, we gain greater explanatory benefits from embracing this multi-levelled approach, because it allows us access to explaining phenomena without appeal to a single standard, which as we saw lead to controversy.

Moreover, the multi-levelled approach provides a way of specifying how cognitive functions can be instantiated, a feature the previous approaches lacked. On the current approach, in order to understand how cognitive functions are realized by a system, and thus check for equivalence, we would need to specify the appropriate immediate successor transition functions – that is, the kinds of structures/states

---

* It is worthwhile to note that nothing in this approach presupposes an intracranial or transcranial view of cognition. It is a purely theoretical tool for conceiving of ways of explaining a system, the realizing basis for that system could be neuronal or otherwise. This means that the objection cannot be raised that I beg the question against either Clark or Adams & Aizawa by presupposing what processes can or cannot count as cognitive.

that serve as representations and the computational processes by which the representations are manipulated. Without specifying the lower lying algorithms understanding the higher cognitive functions for a system becomes problematic (Horgan and Tienson 1993, pp. 164-5). In other words, unless we specify the lower-level algorithms and physical implementations involved in a cognitive process, it becomes difficult to understand the higher-level cognitive functions, e.g. memory. It is primary on this explanatory approach to be able to account for phenomena at all levels of description. All I mean to show with these considerations is that in taking our lead from Marr we can articulate a detailed account of cognition that can incorporate both of what Adams and Aizawa and Clark want to stress is cognitively important. Let me cash this out by applying the approach to Adams and Aizawa's and Clark's accounts.[*]

Adams and Aizawa appealed to the "underlying mechanisms" in their refutations of equivalence arguments. They claimed that these kinds of processes distinguish cognitive from non-cognitive phenomena, because they established a mark of the cognitive. In so doing Adams and Aizawa seem to be drawing our attention to the algorithmic and implementational/computational level distinction. However, they oscillated between the implementational and algorithmic levels in confusing ways. They cited both implementational (neural processes) and algorithmic (recency effects, depth processing etc.) in their refutation of Clark. More problematically, however, is that they built their account of the mark of cognitive only on those two levels and not the third – that is, they made a form of type-identification between cognition and the lower levels of description (implementational and algorithmic).[†] This is why Adam and Aizawa

---

[*] This tripartite distinction is sufficiently more complex then is presented here. However, for my purposes only the general outline needs to sketched. For a more detailed treatment see Horgan and Tienson (1991) or Marr (1977).

[†] This is a point Clark also notes in his response to Adams and Aizawa (Clark 2010, pp. 90-96).

could not address Clark's computational level functional criterion, and so either begged the question or missed the mark in targeting Clark's examples of the CEA.

Clark's arguments, on the other hand, operated only at the computational level. Clark characterized an explanation of a cognitive system in terms of what it does or what functions it computes, e.g. information for memory processing and the resultant output behaviour. This explains why Clark put such a strong emphasis on "functional poise". His argument is predicated on the assumption that if there is relevantly similar information processing and output behaviour then this guarantees cognitive equivalence, and so by extension transcranial cognitive extension. This explains why unless it had already be conceded that the middle and bottom levels constituted cognitive systems, as Adams and Aizawa claimed, Clark's criterion went untouched by their objections.

Nevertheless, even conceding that Adams and Aizawa's case falls short against Clark what is noteworthy in their objections is that they attempt to draw a distinction between the levels of description, i.e. the implementational and algorithmic level on the one hand and computational on the other. This approach, I think, orients us in the right direction to address Clark's CEA.

5.0 THE CEA RE-EXAMINED

Adams and Aizawa insisted that Clark's instances of CEA failed because they did not take the mark of the cognitive into account, but as shown this objection is deeply problematic. However, we do not have to make such problematic claims about what cognition is in order to challenge Clark. Rather, we can propose that the reason Clark's account fails is that it is *incomplete*. That is, insofar as Clark's CEA does not include the algorithmic level of analysis it cannot come to any conclusions regarding extended cognition. Remember that the

isomorphic functional argument moved from functional processing equivalence, in terms of guiding intelligent behaviour, to cognitive equivalence. However, if the need to account for the different levels of description is taken seriously, then the argument needs to be expanded to the following: i) For all x, if x is equivalent to a type of algorithmic/computational information processing, then x is a cognitive process; ii) There exists an x such that x is a type of algorithmic/computational process; and iii) Therefore, x is a cognitive process. Formally expressed it would read:

i) $\forall x\,(AP(x) \rightarrow CP(x))$
ii) $\exists x\,AE(x)$
iii) $\exists x\,CP(x)$

According to this new version of the CEA, the cases Clark points to fail to meet the algorithmic criterion. Thus, they cannot be genuine cases of cognitive extension. However, before I show how Clark's examples fail I should comment briefly on the why this new criterion is relevant.

The reason cognitive equivalence must come at the algorithmic level – along with the computational and implementational levels – is that at the top level too many phenomena are allowed into a cognitive system (Adams and Aizawa's worry against Clark), while at the bottom level what counts as cognitive is restricted to humans, a sort of neural chauvinism (Clark's worry about Adams and Aizawa's account) (Clark 2010 p. 140). It is only at the middle level of algorithms can cognition be properly described. This is because the algorithmic level of description is the most appropriate level for characterizing the abstract system of functional/organizational features that constitute human cognition, and it also provides a transitive link between physical implementational and higher level functions, that is, it can explain how a neural basis can progressively realize higher cognitive functions through the intermediate algorithmic level (Horgan and Tienson 1993, p. 166). These considerations need further argumen-

tation. Nonetheless, the crucial point which emerges is that if we countenance the algorithmic criterion seriously then we can see that Clark's instantiations of the CEA fail not because they are not out and out cognitive but because they are only one part of the proper account. The other part Adams and Aizawa began to indicate when they drew attention to distinguishing cognition at the middle algorithmic and bottom physical implementational levels.

In light of this, how do Clark's examples of cognitive equivalence fair? In the Inga-Otto case the cognitive equivalence is presumably present at the computational or functional level. However, the data-structures that support those cognitive functions are clearly distinct. Otto's cognitive function would be the result of perhaps perceptual and short term memory algorithms while Inga's would be strictly long term memory – this is what Adams and Aizawa seem to be driving at in appealing to underlying mechanisms. These algorithms transitively realize cognitive functions – this means that without middle level equivalence no move to the computational equivalence could be permitted. Hence, Otto and Inga might have the top level equivalence but they fail to be equivalent at the algorithmic level and therefore fail to be legitimate cases of extended cognition.

The same story goes for Clark's second case of the CEA. Granted there may be cases of cognitive equivalence between (1) and (3), in that they may share similar data-structures for supporting mental representations. Nevertheless, the algorithmic criterion, as we have now spelled it out, cleaves (1) and (3) from (2). This is because the cognitive function of mental rotation present in (1) and (3) will be supported in both instances by the same algorithms, neurally realized. However, in (2) the rotation will be carried out by computational algorithms of a different sort – presumably because they are done *in* the computer and not *in* an individual's brain. So, again, the cognitive equivalence cannot be established. In other words, the algorithmic criterion is sharp enough to separate (1) and (3) from (2) such that cognitive equivalence cannot follow the way Clark desires.

If the hypothesis on offer is correct then there are extremely narrow cases of cognitive equivalence, and certainly not equivalence in the cases Clark points to. I think Adams and Aizawa would favour this conclusion because it plausibly preserves the intracranial location of cognitive processing against the present examples. But, the hypothesis also leaves open the possibility that cognitive equivalence spanning the brain, body, and environment at large could be found empirically by phenomena which meet the appropriate description across all three levels, however, this is very unlikely given the restrictions. Nevertheless, we leave it to cognitive psychologists, as Adams and Aizawa recommend, how we should look for such equivalences. Conversely, Clark could be somewhat satisfied with my assessment because it preserves much of what is good about using a functional intelligent behaviour supporting criterion for identifying cognitive processes, insofar as the computational criterion remains salient in my revised version of the CEA. The difference is that my revised CEA no longer makes Clark's computational or functional criterion (Clark's insistence on intelligent behaviour supporting processes) necessary and sufficient to establish cognitive equivalence and thus cognitive extension. In short, cognitive extension can be established only in cases that meet the computational *and* algorithmic criteria, but as it stands Clark has not pointed out any such equivalences with his examples.

6.0 CONCLUSION

The true shortcoming of Clark's CEA examples and Adams & Aizawa's objections lie in their 'incompleteness'. Clark's examples could not establish their intended HEC conclusion because Clark did not include the algorithmic criterion in his version of the CEA, while Adams and Aizawa arguments were either question begging or off the mark against Clark because they identified the cognitive at only the algorithmic and implementational levels and did not take

the computational level of description into account. It is only when all three level of analysis are taken into consideration can we give a full explanation to the range of phenomena which might be plausibly considered cognitively equivalent. It seems, then, that the CEA and the mark of the cognitive, at least as presented by Clark and Adams & Aizawa make poor bedfellows.

REFERENCES

Adams, F., and Aizawa, K. 2001. The Bounds of Cognition. *Philosophical Psychology*, 14 (1), pp. 43-64.

Adams, F., and Aizawa, K. 2010. *The Bounds of Cognition*. Oxford: Wiley-Blackwell.

Clark, A. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. 2010. 'Coupling, Constitution, and the Cognitive Mind'. In: M. Richard, ed. 2010. *The Extended Mind*. Cambridge: MIT Press, pp. 81-97.

Horgan, T., and Tienson, J. 1993. 'Nonclassical Cognitive Science'. In: C. Hookway, ed., D. Peterson, ed. 1993. *Philosophy and Cognitive Science*. Cambridge: Cambridge University Press, pp. 159-87.

Menary, R. 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.

Noë, A. 2004. *Action in Perception.* Cambridge: MIT Press.

Tackling the Regresses: Intentionality and Normativity in Brandom's *Making it Explicit*

EDWARD BROOK
Queen's University

ABSTRACT

Published in 1994, Robert Brandom's Making it Explicit *is an immense treatise on the philosophy of language wherein the author lays bare his unorthodox approach. In this paper I focus on one aspect of Brandom's program – namely, his normative pragmatic theory of meaning – and argue that it is faced with a double regress: the intentionality-regress and the normativity-regress. By contrasting Brandom with Dennett and Searle, however, I show why Brandom's efforts to stop these regresses are successful. While the normative and the intentional are intertwined in Brandom's writing, my strategy here is to separate the regresses and deal with them one at a time. The crux of my demonstration comes at the end of the paper where I explore Brandom's notion of the 'outside' interpreter. This notion enables him to end the normativity regress without non-normative reduction and thus secures his effort to end the intentionality regress, which depends upon the community's normatively significant intentional practices.*

INTRODUCTION

We might say that Robert Brandom's project in *Making it Explicit* involves exploring and explaining the relationship between thinking and talking about things. Of course, this would be a great and unfortunate understatement: Brandom's project is far more ambitious. As Weiss and Wanderer (2010, p. 2) explain, "Brandom aims to construct a self-standing theoretical structure, a unified philosophical system that reveals what it is to say or think something, including one that provides the resources required to state the very system it-

self".

Appreciating the brilliance of Brandom's project in *Making it Explicit* accordingly requires two things. First, one must not lose sight of the forest by focusing too narrowly on trees: crucially, the bigger picture in the book informs and unifies the book's finer details. Second, one must delineate the scope of analysis: the project in *Making it Explicit* is simply too vast to deal with all in one go. In writing this paper, I've tried to keep both of these points in mind. First, for the most part, I've limited the focus in this paper to an analysis of Brandom's normative pragmatic theory of meaning (and thus leaving entire critical chapters untouched). Second, in spite of this, I've dealt with criticisms by placing Brandom's project within the larger context of *Making it Explicit*.

Brandom's approach to language in *Making it Explicit* is unorthodox, and it challenges the way many contemporary philosophers have thought and continue to think about meaning and language. The purpose of this paper, however, is not to champion Brandom's approach directly, but rather to defend it from some ostensibly damaging criticisms. Specifically, this paper is concerned with the fact that Brandom's program is threatened by two troubling regresses: namely, the intentionality-regress and the normativity-regress. My goal in this paper is to show why Brandom's efforts to stop these regresses are successful.

BRANDOM'S PROJECT AND THE TWO REGRESSES

Brandom explains meaning in terms of use, semantics in terms of pragmatics. He believes that the conceptual content of linguistic expressions is borrowed or derived from the thoughts and beliefs of the language users who interpret them. In this sense, he believes the act of assert*ing* takes priority over what is assert*ed*. Brandom's account of pragmatics is also fundamentally *normative*. He understands the conceptual content of linguistic expressions in terms of the *proprie-*

ties that govern language use. Hence, there are two different forms of regress-argument that might be used against Brandom's normative pragmatic theory of meaning.

First, insofar as Brandom adopts the Dennettian picture of intentionality – according to which having intentional states is a matter of being *ascribed* intentional states – one might argue that his account involves an 'intentionality-regress'. Brandom needs to explain where the ascribers of intentionality get their intentional states, and where the *ascribers* of ascribers get their intentional states, and so on. Second, insofar as Brandom believes the norms governing language use are instituted by the practical attitudes of those who acknowledge them, one might argue that this account involves a 'normativity-regress'. Brandom needs to explain how norms govern the activity of instituting norms, and how norms govern the activity of *governing the activity* of instituting norms, and so on. In short, Brandom must either ground intentionality and normativity in something fundamental, or explain why – counter to our intuitions – such grounding is not required.

Brandom does a little of both: he provides an account of original intentionality and normativity using only the resources available in *Making it Explicit*, but does so in an unconventional manner.[*] In response to the first regress challenge, Brandom argues that the linguistic community is the ultimate source of intentionality. "Only communities," he writes, "can be interpreted as having original intentionality" (1994, p. 61). This *social* grounding of intentionality challenges the view of intentionality defended by Searle, who believes that making sense of derivative intentionality requires that intentional interpreters possess intentionality intrinsically. This approach

---

[*] While 'original intentionality' is a term used by Dennett and Brandom, 'original normativity' is not. I use the latter term to talk about the origin of norms, which is distinct from the *objectivity* of norms insofar as one might locate original normativity in the community and have nothing to say about the community being objectively correct or incorrect.

also challenges Dennett's view of intentionality insofar as Brandom stops short of explaining where the linguistic community derives *its* intentionality (or so Dennett argues.) In contrast, Dennett believes original intentionality must be grounded naturalistically in the process of evolution (2010). He writes: "the only 'original' intentionality anywhere is the mere *as-if* intentionality of the process of natural selection viewed from the intentional stance" (2010, p. 52). This raises the following questions: How do Brandom and Dennett differ from Searle? How does this enable them to avoid positing intrinsic intentionality in interpreters? Moreover, why can Brandom stop short of Dennett's evolutionary naturalism? Why is this latter question relevant for Dennett, but not particularly crucial for Brandom?

In response to the second regress challenge, Brandom argues that the normative status of conceptual content is, similarly, grounded in the community. What is meant by 'community' in the normative case, however, differs from the intentional case mentioned above: according to Brandom, normative statuses are instituted by practical attitudes while intentional contents are *conferred* by social practices in which those statuses and attitudes play a role. The objectivity of the norms governing language use, for Brandom, arises from the inherently social distinction between acknowledging a commitment and attributing a commitment to another person. Brandom's approach to grounding norms deviates from Dennett's picture insofar as Dennett explains the normativity of conceptual content naturalistically in terms of predictive success. Brandom's approach also differs from normative pragmatic theories of meaning wherein norms are grounded socially, but in terms of the proprieties governing a public language. This raises the following questions: How can Brandom ground normativity in a social distinction between attitudes? How does his strategy differ from Dennett's? Again, why can Brandom stop short of Dennett's naturalism? Lastly, what is Brandom doing such that his approach differs from philosophers who rely upon a public language to ground normativity?

This paper explores Brandom's normative pragmatic theory of meaning and in particular, his strategies for dealing with the respective regress challenges. After laying out the finer details of Brandom's approach to intentionality, I argue that his method for ending the intentionality-regress is successful. I suggest that Searle's claim that interpreters have intrinsic intentionality is problematic and I show how the Dennettian model avoids having to make this kind of claim. I also suggest that Dennett's criticism of Brandom – namely, that he needs to explain where communities get intentionality – involves a misunderstanding of Brandom's central project. The community, for Brandom, is not a singular entity. While his language in *Making it Explicit* may appear to commit Brandom to such a view, appreciating his work in its entirety removes the temptation to read Brandom as Dennett does. Subsequently, I argue that Brandom's method for stopping the normativity regress is successful. The important thing to show here is that a violation of conceptual norms is not merely a *faux pas*, but as Dennett puts it, "a mistake that really matters". Brandom's defence of the *I-Thou* sociality in which attributions and acknowledgements of commitment and entitlement give rise to a distinctively perspectival form of intersubjectivity provides the objective basis that conceptual norms need.

DENNETTIAN INTENTIONALITY AND BRANDOM'S PROJECT

Brandom's theory of meaning begins with *sapience*: he undertakes to explain conceptual content in terms of the kind of discursive, linguistic creatures we are. This route into meaning can be contrasted with a project that begins with *sentience*. Sentience is what we share with non-verbal, non-linguistic animals – it is a form of awareness. Sapience is a different matter insofar as it involves specifically propositional conceptual contents and specifically normative features. As Brandom writes, "we are the makers and takers of reasons, seekers and speakers of truth" (1994, p. 6). Accordingly, Brandom begins by

looking at the kind of intentionality we have as linguistic, concept-using, and reason-respecting creatures – that is, intentionality in the sense of propositionally contentful attitudes. Brandom acknowledges, of course, that his project must be capable of saying something about the beliefs and desires of non- or pre-linguistic animals, or else risk "being beastly to the beasts" (1994, p. 7). The Dennettian picture of intentionality allows Brandom to give explanatory priority to the kind of intentionality we possess and still be in a position to explain the intentional states of merely sentient intentional systems.

Dennett understands intentionality in terms of the ascription of intentional states: being an intentional system, for Dennett, involves being taken as one. When we play chess with a computer, we ascribe it certain beliefs, desires, and other intentional states (for example, the desire to win). The computer's intentionality is wholly derivative on our own. This basic idea informs Dennett's approach. Dennett distinguishes between first-order intentional systems, whose behaviour is interpretable by ascribing beliefs, desires etc., and second-order intentional systems, who are capable of ascribing intentional states to others and to themselves (2010, p. 49). Brandom adopts Dennett's ascriptivist/stance approach. He refers to first-order intentional systems as *simple intentional systems* and second-order intentional systems as *interpreting intentional systems* (1994, p. 59).[*] Insofar as Brandom, like Dennett, thinks that simple intentional systems derive intentionality from interpreting intentional systems, he is faced with the following question: where do second order intentional systems get their intentionality from?

DENNETTIAN INTENTIONALITY AND THE FIRST REGRESS

Brandom is aware of this question and the regress challenge it poses

---

[*] The idea here is that an "interpreting intentional system" is an entity capable of *interpreting* its own behaviour and the behaviour of other systems.

for the Dennettian 'stance' stance.* He writes: "On the pain of infinite regress, it seems necessary to distinguish the *derivative intentionality* such merely interpretable items display from the *original intentionality* their interpreters display" (1994, p. 60). One manner of dealing with the threat of an infinite regress, favoured by Searle, is to posit *intrinsic* intentionality in interpreting intentional systems (1983). Searle writes: "the false premise in the argument in short is the one that says that in order for there to be a representation there must be some agent who *uses* some entity as representation. This is true of pictures and sentences, i.e., of derived Intentionality, but not of Intentional states" (1983, p. 22). For Searle, it is simply a brute fact that interpreting systems (i.e. humans) possess original intentionality. The difficulty with this view is that it injects mystery into the problem at hand. What does "intrinsic" intentionality mean? Why would we treat intentionally interpretable chess-playing computers differently from intentionally interpretable language-speaking concept users? The problem with Searle's approach is that instead of explaining how intentionality works, it simply puts the question aside and moves on. As Dennett writes, "Searle's vision of original intentionality is ultimately incoherent; it begins with an unacceptably marvellous and unexplained prime mover" (2010, p. 52).

Unlike Searle, Dennett and Brandom locate original intentionality outside of particular interpreting systems. According to Dennett, the intentionality of interpreting systems is just as derived as posters or chess-playing computers. Roughly, he argues that we get our intentionality from our selfish genes, which get their intentionality, indirectly, from the process of natural selection. Dennett believes that the 'rationalistic calculus' of interpreting and predicting is an

---

* The *'stance' stance* is a term of art used by Brandom to define this controversial position articulated by Dennett: namely, that there is no room for a distinction between being an intentional system and being appropriately treated as one. In other words, treating something as an intentional system – for both Dennett and Brandom – is prior, in terms of explanation, to that of being an intentional system (Brandom 1994, p. 57).

instrumentalistic method that has evolved simply because it works (1987, p. 48). Intentional systems, Dennett argues, are those things for which it is predictively useful – and thus evolutionarily beneficial – to adopt the intentional stance toward (Brandom 1994, p. 57). Hence, Dennett says that "*our own* intentionality is just as derived as that of our shopping lists and other meaningful artifacts, and derives from none other than Mother Nature, the process of evolution by natural selection" (2010, p. 51). Predictive utility is understood in terms of survival and thus natural selection. As such, original intentionality is ultimately located in nature. The fact that evolution lacks beliefs, desires, and purposes does not bother Dennett, however. As he views it, this "permits us to end the regress of ulterior purposes with a whimper, not a bang" (2010, p. 51). What he means by this is that eventually the need for intentional explanation will run out because "Mother Nature" is not a system, but rather in Dennett's evolutionary sense, a non-intentional process. In contrast, Brandom argues that original intentionality is located in the community. Thus, whereas Dennett thinks that interpreting intentional systems derive their intentionality naturalistically, Brandom thinks that these higher intentional systems get their intentionality *from each other* (2010, p. 51). "On this line," writes Brandom, "only communities, not individuals, can be interpreted as having original intentionality" (1994, p. 61).

Brandom's account of *how* we confer intentional states and *where* original intentionality is ultimately located (i.e. the community) is difficult to separate from his discussion about normativity – and thus difficult to separate from the question of a normative regress. Nevertheless these two aspects of Brandom's project are intelligible when taken apart. "The idea of normative statuses as *instituted* by practical attitudes," Brandom writes, "should be distinguished from the idea of their intentional contents as *conferred* by the social practices in which those statuses and attitudes play a role" (1994, p. 64). Consequently, before looking at the important role normativity has

to play on Brandom's account, we can consider his account of intentionality independently to get a better sense of how he will address the regress problem.

BRANDOM'S TAKE ON DENNETTIAN INTENTIONALITY

First, Brandom's account of intentionality is fundamentally pragmatic. For him, "talk of what is expressed is intelligible only in the context of talk of the activity of grasping what is expressed" (1994, p. 73). Second, Brandom understands the activity of grasping what is expressed in terms of inference and the incompatibility of inference – that is, he proposes an *inferential* demarcation of the conceptual (1994, p. 89). This inferential demarcation, of course, is understood pragmatically in terms of *inferring*. Third, the inferential proprieties governing intentional content is explained as a pragmatic theory *implicit* in the discursive practice of a community (1994, p. 133). For Brandom, certain features of the practices that involve conferring conceptual content on things and people are what make the community the locus of original intentionality. The most important feature is the *normative* significance that governs the ascriptions of commitment and entitlement that make up these practices. This will become clearer below when I address Brandom's discussion of an 'outside' interpreter conferring original intentionality on a community. It is *not* the case that the community's original intentionality is derivative upon or borrowed from the interpreter (by definition it cannot be), but rather that the features of the inferential, norm-governed ascriptive practices in that community are sufficient for original intentionality. Thus, for now, it will suffice to say that if asked "where do interpreting systems get their intentionality from?" Brandom would reply "they get it from each other".

A NORMATIVE END TO THE INTENTIONALITY REGRESS

It should be clear, then, that normativity is important to both Dennett and Brandom's strategies for ending the intentionality regress. As Dennett writes, "normativity [...] alone can give traction to a regress stopper that we need" (2010, p. 50). For Brandom, "Insofar as their intentionality is derivative – because the normative significance of their states is instituted by the attitudes adopted toward them – their intentionality derives from each other, not from outside the community" (1994, p. 61). For Dennett, intentional explanation based on predictive utility is fundamentally normative insofar as it involves a substantive rationality assumption: "A system's behaviour will consist of those acts that *it would be rational* for an agent with those beliefs and desires to perform" (1987, p. 49). Thus normativity, for Dennett, is asserted into his naturalistic approach by assuming that intentional systems are rational.

For Brandom, the community exemplifies original intentionality insofar as the members' practical attitudes institute normative statuses and confer intentional content on them. For Dennett, the rationality assumption introduces the notion that an intentional system *ought* to have certain beliefs and desires. In short, normativity is what makes each of these responses possible. In either case, the introduction of normativity buttresses the explanation of original intentionality, and thus stops the regress without an appeal to intrinsic intentionality.

However, relying on normativity to ground intentionality gives rise to a second infinite regress. When we understand the institution of norms as the attribution of normative statuses by interpreting systems – as Brandom and Dennett both do – then we must be willing to ground norms in something. As Dennett writes, "They aren't just prejudices or fashions that have unaccountably caught on" (2010, p. 50). If norms are instituted by our practical attitudes and the practice of instituting norms can be done correctly or incorrectly, then there must be another level of norm-governed practical attitudes in-

stituting norms, and so on. As such, Brandom and Dennett must say something about the normative significance of intentional explanation in order to avoid the second potential regress.

"NORMS ALL THE WAY DOWN"

Both Brandom and Dennett treat the normative significance of intentional states as instituted by the attitudes of interpreters (1994, p. 58). The point where they diverge is instructive for understanding what makes Brandom's version of normative pragmatics so unique. In *The Intentional Stance*, Dennett argues that the norms governing the attitudes of interpreting systems can be explained in non-normative terms (1987). Dennett's claim that we adopt the intentional stance based on its predictive success reduces talk of propriety to talk about the behaviour of rational animals. The substantial rationality assumption, which introduces the fact of how a system *ought* to behave, can be explained in non-normative terms – for example, in terms of natural selection and competing survival mechanisms. Brandom, however, rejects this strategy. He argues that we cannot naturalistically reduce the normative to the non-normative. On his view, "it's norms all the way down" (1994, p. 44). As Gideon Rosen (1997, p. 163) observes, "Brandom defends this surprising view [...] maintaining that the normative idiom cannot be made intelligible in *any* more basic terms. In the conceptual order, the normative is strictly prior to the intentional; but nothing is prior to the normative itself". Accordingly, there are two things an account like Brandom's must provide: first, an explanation of how this "norms all the way down" picture works – that is, how the norms governing a community are instituted by the community; and second, an explanation of how there can be *objective* norms – that is, how it can be possible that every member of the community is wrong. Brandom's strategy involves the image of a community where members view each other (and themselves) as being committed and entitled by the assertions

they make. This approach a) explains the institution of norms in terms of practical attitudes; b) takes norms to be *implicit* in discursive practice; and c) is a fundamentally social account of normativity.

NORMATIVITY: AVOIDING REGULISM AND REGULARISM

Brandom needs a solid story about how the norms of a community can govern its discursive practices – that is, for example, why it is wrong to say "I like your dog" when you actually mean "Can I borrow your lawnmower?" The problem, however, is that there have been similar attempts before, and for the most part, they have been unsatisfactory. Accordingly, for Brandom's project to succeed, he must weave around the problematic approaches.

For Brandom, norms are instituted by normative attitudes. He takes this idea from Kant, who claimed that we are not only subject to norms but also sensitive to them (1994, p. 33). Brandom, however, rejects Kant's corresponding claim that norms are rules. For Kant, acts are subject to normative assessment insofar as they are governed by explicit prescriptions or prohibitions (1994, p. 19). The problem with this view – which Brandom terms 'Regulism' – was made evident by Wittgenstein in his rule-regress argument. When norms are conceived of as rules they only have normative significance insofar as they are *correctly* applied. Accordingly, there must be explicit rules governing the application of rules, and rules governing the application of those rules, and so on (1994, p. 20). Wittgenstein argued that it was *not* the case that there are 'rules all the way down' but rather that the regress terminates in some kind of practical primitive – that is to say, the regress will end when we focus on the practices of the group or community in question. Hence, the conclusion of the regress argument, Brandom writes, "is that there is a need for a *pragmatist* conception of norms – a notion of primitive correctness of performance *implicit* in *practice* that precede and are presupposed by their explicit formulation in *rules* and *principles*" (1994, p. 21). Prag-

matism about normativity will be a crucial aspect of Brandom's effort to ground the norms governing discursive practice. In his view, the norms governing what one should think and not think, or say and not say given the circumstances, are proprieties implicit in our practice, in what we do. Thus, by understanding normativity pragmatically, Brandom has already lessened the blow of the normativity-regress. It is no longer as vicious as the regress that beleaguered Kant. Nevertheless, there is still a question about the norms that govern the correctness or incorrectness of the normative attitude and this, it seems, cannot be explained by appealing to the normative attitudes of others without giving rise to a regress.[*]

An alternative to regulism about norms, which Brandom considers but rejects, is regularism about norms. Regularism is the view that talk about 'norms implicit in practice' is just talk about regularities of behaviour (1994, p. 27). The regularist account avoids the regulist regress, but it nevertheless encounters serious challenge. First, Brandom argues that it removes the contrast between treating a performance as subject to normative assessment and treating it as subject to physical (and thus non-normative) laws. We cannot build a story about norms implicit in practice by appealing to regularities in behaviour because this eliminates the possibility of mistakes: whatever the individual system or community of systems does will always be normatively what it ought to do (1994, p. 27). It should be clear, then, why Dennett introduces his 'rationality assumption'. On his naturalistic approach, he needs some way of bringing normativity *back* into the picture insofar as he uses the behaviour of intentional systems to ground the intentionality regress. We are not merely ascribing normatively significant intentional states via predictive utility, but ascribing them to *rational* creatures that *ought* to have certain beliefs and desires as such. The main problem with regularism, however, is

---

[*] By "normative attitude" I simply mean the attitude of taking or treating applications of concepts as correct or incorrect. At this point, by rejecting regulism, practical "normative attitudes" are the new primitives.

that it allows for gerrymandering.* The gerrymandering objection against regularism is premised on the fact that an irregularity can only be specified with reference to a previously identified regularity and therefore depending on which regularity is specified, something will count as an irregularity in some cases and a regularity in others. Brandom writes: "A performance can be denominated 'irregular' only with respect to a specified regularity, not regularity *tout court*. Any further performance will count as regular with respect to the patterns exhibited by the original set and as irregular with respect to others" (1994, p. 28). Accordingly, regularity theories must either smuggle in some normative fact about which regularity should be specified or rescind the claim to have offered a genuinely normative picture.

Thus, the problem with regularity theories, according to Brandom, is that they fail to take seriously Kant's distinction between acting according to a rule and acting according to one's conception of the rule. As a consequence, regularity theories remove the possibility of mistakes. Even if regularity theorists, such as Haugeland, explain norms in terms of positive and negative sanctions, the gerrymandering problem persists because sanctioning can be done correctly or incorrectly, and if the norms governing sanctioning are mere regularities of enforcement, then it becomes a question of being *actually* punished as opposed to being correctly punished (1994, p. 36). Thus, regularity theories have two problems: 1) they collapse the normative and the non-normative together, thus violating Kant's important insight that acting according to a rule is different than acting according to one's conception of a rule; and 2) they raise the problem of gerrymandering – the problem of specifying which regularity is the regularity in question.

---

* The term 'gerrymandering' refers to the process of re-arranging electoral districts so as to ensure a victory at the polls. The same idea applies here: one can always gerrymander the regularity to include a particular irregularity (thus removing the irregularity.) Consider the pattern XyXyX. It seems that a Z would be an irregularity (XyXyXZ.) However, one can gerrymander the pattern like this: XyXyXZXyXyXZ.

AVOIDING COMMUNAL ASSESSMENT REGULARISM TOO

Brandom's solution is to focus on the *social* institution of norms. His understanding of social, however, is different from that of regularity theorists who would explain the institution of norms in terms of communal assessment. On this latter approach, individual violations of norms are made possible by the idea that individual performances are assessed and responded to by the community. As Brandom writes, "the community, unlike the individual, need not be counted as having taken up a practical attitude regarding the propriety of the performance just in virtue of that performance's having been produced by one of its members" (1994, p. 37). In other words, communal regularity theories appear to avoid the gerrymandering problem because a) the focus is on assessments rather than behavioural regularities, and b) because individual members can act incorrectly without the community having to count as condoning these actions.

Communal regularity theories are still problematic, however. According to Brandom, the problem with the 'I-We' sociality, is that: 1) in spite of being a *complex* regularity theory, it still must smuggle in normative language despite being purportedly nonnormative; and 2) the idea of communal performances and assessments relies on a fictional 'community' a (1994, p. 38). The first problem is one that we have already seen: regularity theories of norms are subject to two sorts of objections – the gerrymandering challenge, and the fact that they neither do justice to the essential features of normative statuses, nor covertly appeal to normative notions.

The communal assessment approach may avoid the gerrymandering challenge, but it falls come the latter issues. This is because being a community member brings with it normative notions and specifying what it is for something to count as communal attitude smuggles in normativity (1994, p. 41). As Brandom writes, "Regularity theories, of whatever stripe, are (to adapt a phrase of Dretske's) attempts to bake a normative cake with nonnormative ingredients" (1994, p. 43).

Inevitably, they end up smuggling in normativity without explaining how or why. The second problem with the notion of communal assessment is that it introduces the notion of a mythical community wherein the *community* somehow acts in unison, assessing and attributing individual community members. In addition to being difficult to grasp, this concept is problematic insofar as it risks making the community incorrigible, that is to say, it makes it impossible for the community as a whole, to be wrong. While there are clearly some norms like this – for example, the appropriate greeting gesture in community X will be whatever the members of community X have decided, through tradition or habit, the appropriate gesture to be – there is a sense in which the objective basis for normativity must include the community (1994, p. 53). It must be possible for the entire community to be wrong about some conceptual norm.

For example, suppose that everyone in the community believes that the nearest water source is 2 kilometre to the north of them. When a member of the community says, "I'm going to the nearest water source" they mean that water source 2 kilometres away. Other members of the community attribute this meaning to their statements and will sanction or approve of the statement depending on the context in which it is made. If someone says "I'm going to the nearest water source" as they lie down for a nap, they will be sanctioned; that is, the other people will give that person a funny look and not understand what he meant. Alternatively, if someone says, "I'm going to the nearest water source" as they leave the community with a bucket in hand, people will approve of how the conceptual content was used. But it must be possible for the whole community to be wrong. Suppose that the nearest water source was actually 1 kilometre away but no one in the community knew it was there. In this scenario, the entire community is wrong by sanctioning and approving of the statement "I'm going to the nearest water source" in the manner they do now.

Consequently, some have argued that the objectivity of norms (in the sense that the community might be wrong) is something that we

have to forego if we insist that normative statuses are instituted by social practices (1994, p. 54). Brandom disagrees with this conclusion; he believes that we can have objective norms while nonetheless maintaining that they are the products of practical attitudes implicit in a fundamentally social practice.

Brandom's project in *Making it Explicit* is to articulate a social practice that governs the institution of the norms underlying propositional attitudes but does not depend on communal assessment. Accordingly, Brandom's account of the objectivity of conceptual norms through the practical attitudes of community members appeals to the social distinction between acknowledging a commitment oneself and *attributing* a commitment to another (1994, p. 55). Brandom does this by privileging an 'I-thou' sociality rather than an 'I-we' sociality. The basic notion here is "the relation between an audience that is attributing commitments and thereby keeping score and a speaker who is undertaking commitments, on whom the score is being kept" (1994, p. 508). Brandom articulates this approach succinctly in the final pages of Chapter Eight of *Making it Explicit*. The objectivity of conceptual norms is presented as a feature of the formal structure of discursive intersubjectivity – that is, intersubjectivity understood in the perspectival 'I-Thou' sense. The problem with understanding intersubjectivity in the collective, 'I-We' sense is that the privileged position afforded to communal assessment means that, what the community *takes* to be correct *is* correct (1994, p. 599). In contrast, "according to the 'I-Thou' construal of intersubjectivity, each perspective is at most *locally* privileged...What is shared by all discursive perspectives is *that* there is a difference between what is objectively correct in the way of concept application and what is merely taken to be so, not *what* it is – the structure, not the content" (1994, p. 600). Thus the interaction of community members with each other that gives rise to objectivity – an interaction which involves attributing and acknowledging inferential commitments and entitlements – is conceived by Brandom structurally rather than materially.

It will be helpful to recapitulate Brandom's story about the origin and nature of norms just presented in order to appreciate how this provides him with a solution to the normativity-regress identified above. First, the norms governing conceptually structured activity are instituted by our practical attitudes. The normative significance of intentional states, Brandom writes, is instituted by the practical attitudes of interpreting intentional systems. This stems from Kant's observation that following a rule must be distinct from conceiving of a rule if the act is to be truly normative. Second, these norms are implicit in our practice. Although they may manifest in explicit prohibitions or prescriptions, norms must be implicit in our practice as shown by Wittgenstein's rule regress argument. They cannot be implicit as understood in terms of behavioural regularities, however. Both the gerrymandering challenge and the threat of collapsing normative statuses into normative assessments (which means either losing or smuggling in normativity) demonstrate the inadequacy of regularity theories. Third, implicit norms are inherently social. They must be social insofar as the perspectival, 'I-Thou' symmetry of discursive attitudes funds the intersubjectivity needed for the objectivity of norms. Thus, even though "we discursive beings are creatures of norms" and "norms are in some sense creatures of ours" (1994, p. 626), the norms governing conceptual content can be objective in the sense that the entire community *could* be wrong.

NORMS ALL THE WAY DOWN?

In the preceding section I suggested that Brandom can, in fact, fund the idea of objective norms by appealing to intersubjectivity understood in an *I-Thou* sociality. Brandom conceives of objectivity as a formal, as opposed to a substantive, characteristic of norms (1994, p. 597). Objectivity, for Brandom, is a structural aspect of a community's discursive practice – not some sort of content that norms either have or do not have. In this picture, the possibility of different

I-Thou perspectives ensures the possibility that a given speaker or group of speakers is wrong.

Nevertheless, there is one last piece to this puzzle – namely, how a community of interpreting intentional systems can be said to participate in genuinely normative practices when the objectivity of conceptual norms is premised on the *I-Thou* relation between particular speakers and listeners. As Danielle Macbeth writes, because "there need be no essentially shared public language, no shared set of norms, even implicit in practice," Brandom's picture "cannot fund any notion of correctness, hence of content (whether objective or not) at all (1994, p. 201). In other words, Brandom's "norms all the way down" picture has some loose ends. This may be why Dennett criticizes Brandom for appealing to "community" as a mysterious primitive. "Community is Brandom's skyhook," Dennett writes, "and he can't have it" (1994, p. 55). Simply put, Brandom's critics will not let him use the concept of community to unite the disparate speakers and scorekeepers under a common normative practice – especially after his disavowal of community assessment regularity theories and his similar rejection of the community as a mythical entity.

Accordingly, it may seem as though Brandom is smuggling normativity back into the picture by asking the reader to fill in the gaps in his story about norms. In fact, it may be that Brandom needs an outside interpreter to ascribe normatively significant intentional states to the community in order for there to be a meaningful notion of correctness. As Macbeth suggests, "what is at issue is whether any content can be given to the idea that I in acting, or indeed you in assessing, do something normatively significant at all" (2010, p. 201). The last piece of the puzzle addresses this very concern. Brandom's explanation of norms does, in fact, involve an appeal to the notion of an outside interpreter. Not only does the interpreter take up the intentional stance towards the community (giving it original intentionality), it also takes up a practical scorekeeping attitude towards the community (ascribing normative significance to those inten-

tional states.) In short, Brandom's outside interpreter *makes explicit* what the community does implicitly and thereby gives traction to his overall project, bringing the normative regress (and thus the intentionality regress) to an end.

THE OUTSIDE INTERPRETER

Brandom acknowledges the challenge of explaining how both original intentionality and normativity can rest in the community – especially on his model wherein intentional states are conferred by the stances interpreting systems take towards each other, and wherein normative statuses are the products of the practical attitudes of discursive agents. The problem of showing how "norms all the way down" can work raises two question. First, what must one be doing in order to count as taking a community to be engaged in implicitly normative practices? That is, what must the *interpreter* be doing when she confers normative status on the community? Second, what is it about the performances, dispositions, and regularities exhibited by a group of sapient creatures that makes it correct or appropriate to adopt that attitude? In other words, what is the community doing so that the interpreter adopts the normative attitude towards it?

Brandom's response to the first question involves a close consideration of the differences between the stance of interpreting intentional systems within the community and the stance of the outside interpreter. Crucially, there is a difference between attributing genuinely linguistic practices to a community and attributing intentional states to simple systems. In the latter case, the "semantic ingredient" is supplied by the interpreter rather than attributed to the system being interpreted (1994, p. 630). In the case of the community, original intentionality is attributed to the community rather than the interpreter. Both cases involve taking a 'stance', but in the latter case, the structural features of the practice are what make it correct or incorrect for the interpreter to attribute original intentionality.

Consequently, we must resist the urge to characterize the difference between the community members and the outside interpreter as stemming from the fact former are attributing normatively significant intentional states to their *own* members while the interpreter is taking the 'stance' stance towards members of some *other* community (1994, p. 638). This characterization overlooks the central tenet of Brandom's program, namely, making *implicit* practices *explicit*. The salient difference between the interpreter and the interpreting community members is that the interpreter, according to Brandom, is "making explicit the broadly inferential properties that are *implicit* in their scorekeeping practices" (1994, p. 639). Accordingly, Brandom removes any internal/external framework, which had seemed to support the interpreter's stance. Moreover, by explicitly keeping discursive score on the members of a community, the interpreter expands the "we-saying" attitude to include those members. As Brandom writes, "the sort of scorekeeping that is – according to the interpreter outside the community – internal to and constitutive of the community being interpreted *comes to coincide with the scorekeeping of the interpreter who attributes discursive practices to the members of that community*" (1994, p. 644; my emphasis).

The second question concerns the origin of norms: specifically how it is that communities end up with practices that are normatively significant at all. This is one question that seems to be motivating Macbeth's criticism. Specifically, the question is concerned with the norms governing the interpreter's adoption of the interpretive stance towards a community. What makes the interpreter's conferral of normatively significant intentional states correct in a given instance? Now this would be a difficult question to answer if we had not already collapsed the external interpretive position of interpreter and the internal scorekeeping practices of the community. The norms governing the 'outside' interpreter's activity are to be explained in terms of *our own practices*. Brandom believes the question becomes much easier to deal with once we see this fact. When is it appropriate

to adopt the discursive scorekeeping stance towards a community? It is appropriate whenever one *can* adopt it. As Brandom writes, "Not just any group of interacting organisms can be made out to be attributing to each other commitments whose inferential and social articulation suffices to confer genuinely propositional contents on their performances" (1994, p. 644). Accordingly, there is little risk that 'undeserving' communities will be attributed original intentionality.

Thus, it is not the case that original intentionality is derived from the interpreter in the same way the intentionality of simple intentional systems is derived from interpreting intentional systems. Rather, the interpreter attributes intentionality on the basis of whether or not the practices of the community members are sufficient to confer normatively significant intentional states on each other. The propriety of adopting the intentional stance is likewise dependent upon whether it is appropriate to extend the characteristically sapient 'we-saying' attitude towards certain organisms. In turn, the act of treating other systems as being "like us" is governed by norms within our own practice. Accordingly, norms originate from our norm-governed practice of instituting norms. As Brandom says, "it's norms all the way down." What the interpreter is doing is making explicit these implicit norm-instituting practices.

At various points throughout *Making it Explicit* Brandom describes his project as one that requires "bootstrapping" or "pulling oneself up by the boots." This analogy is apt because it gets to the heart of what the interpreter is doing by making it explicit. The outside interpreter is attributing to the community of interpreting intentional systems a discursive scorekeeping practice that is fundamentally normative. This outside interpreter, however, is also partaking in this same normative, discursive scorekeeping practice and is therefore bound by the norms implicit in the practice. Hence, in one sense at least, the regress ends by coming around full circle. Or as Brandom writes, "along this expressive dimension, the project eats its own tail…" (1994, p. xx).

CONCLUSION

In conclusion, we can think of Brandom's proposed system as a three-part cycle. First, talking and thinking is described in terms of inferentially articulated norms. The inferential moves in the discursive practice are understood in terms of commitment and entitlement. Second, treating or taking interlocutors as being committed and entitled by their assertions is governed by norms implicit in our practice. Deontic statuses are instituted by the proprieties of scorekeeping through the adoption of deontic attitudes by the community. Third, attributing deontic attitudes – that is, interpreting a community as engaged implicitly in normative practices – is explained by reference to an interpreter who is effectively *making explicit* the implicit scorekeeping practices of the community. The issue of when it is appropriate to interpret a community as exhibiting original intentionality comes around full circle to an issue involving the particular deontic statuses implicit in our practice at the first level of norms (1994, p. 638).

Brandom understands semantics in terms of pragmatics and takes pragmatics to be fundamentally normative. His semantic pragmatism is manifested in the belief that meaning is derivative on sapient creatures like us. Brandom's commitment to *normative* pragmatism is exemplified in the Kantian framework he adopts to explain intentionality. What separates us from reliably responsive thermostats, for example, is that we can conceive of and adhere to the norms governing conceptual content. Brandom's normative pragmatics ultimately finds its form in Dennett's stance-centred approach to intentionality. According to Dennett, intentional states such as beliefs and desires are ascribed to objects and systems by higher intentional systems. Likewise, for Brandom, interpreting intentional systems confer intentional status on the assertions and action of people and things. However, whereas Dennett believes that all intentionality is fundamentally derivative intentionality, and locates as-if *original* intentionality in the process of natural selection, Brandom rejects natural-

istic reduction and locates original intentionality in the community. While they differ on *where* intentionality gets grounded, both Brandom and Dennett agree that the ascription of intentional states by interpreting systems is fundamentally normative. It is this appeal to normativity that enables Dennett and Brandom to ground intentionality and thus avoid the threat of an infinite regress.

Appealing to normativity to explain the conferral of intentional content begets a second regress, however. If norms govern the ascriptions of intentional states then there must be correct or incorrect ways of applying these norms. Brandom avoids the brunt of the regress by understanding the norms in terms of the institution of deontic statuses and the institution of those statuses in terms of the practical attitudes of interpreting intentional systems. There remain two questions, however, regarding the origin of social norms. First, how can proprieties governing the correct or incorrect institution of deontic statuses be instituted *by* the community they govern? And second, what makes these norms objective in the sense that the community they govern could be wrong? Brandom tackles this latter question by explaining norms via an *I-Thou* sociality – in contrast to an *I-We* sociality. By appealing to the perspectival nature of the scorekeeping practice, Brandom is able to fund a structural version of objectivity. He is thus able to explain how everyone in the community could be wrong. Brandom addresses the first question by explaining how an outside interpreter takes up a practical attitude toward the community. As with original intentionality, it is not the case that the normative significance of the community's practices is somehow derivative upon the interpreter (in which case the regress would continue.) Instead, the interpreter is merely making explicit the practices that are taking place implicitly within her own community.

Thus, we can put aside Dennett's original criticism. The problem with Dennett's analysis, as we have seen, is that Brandom *can* make sense of original intentionality (and original normativity) by appealing to an 'outside' interpreter who makes explicit the implicit

intentionality-conferring practices of the community to which she belongs. Dennett asks us to consider the same situation, but with a community of robots. In this case, we would ask "where do the robots get their intentionality?" and expect a substantial answer (1994, p. 52). The disanalogy here is important. The interpreter is *one of the members* of the community in a way in which a 'robot-maker' is not a member of the robot community. Dennett wants a further evolutionary story about how humans got to be in the position they are – that is, capable of attributing intentional states, and of making explicit practices implicit in their community. But Brandom is not concerned with this issue in *Making it Explicit*. As he admits, "The question of how the trick is done…is explicitly put to one side" (2010, p. 306). Brandom begins from sapience and moves forward from there. In other words, he is focused on the trick itself – namely, on the articulation of social norms and the way they govern concept use in the community. Hence, whether Dennett is right about Brandom having to *eventually* provide a naturalistic/evolutionary story about how sapient beings arose and formed communities, it is not relevant to original intentionality. As Brandom writes, explaining the advent of language is a separate, albeit complementary question to his project (2010, p. 308).

Moreover, the criticism that Brandom cannot give a meaningful explanation of correctness involves a misunderstanding of the interpreter's role. The norms governing the interpreter are the same norms that are instituted by the practical attitudes of the community. This is because taking the community to be capable of broadly-inferential discursive scorekeeping the interpreter must extend the 'we-saying' attitude of sapience to include those organisms. Thus, Brandom's program succeeds in spite of being circular and in spite of "bootstrapping itself" into working order. Philosophers tend to be afraid of circular things – but not all circles are bad. To the extent that the interpreter is a member of the community and is making explicit the implicit practices of that community, she is bound by the

very norm-governed practices she is interpreting. Once again, what the discursive scorekeeper within the community does implicitly, the interpreter does *explicitly*. The underlying difference is a distinction between implicit and explicit interpretations (1994, p. 639). It is in this way that Brandom avoids two potentially damaging regresses – one concerning the origins of intentionality and one concerning the origins of normativity.

REFERENCES

Brandom, R. 1994. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.

Brandom, R. 2010. 'Reply to Daniel Dennett'. In: B Weiss & J Wanderer, eds. 2010 *Reading Brandom: on Making it Explicit*. Abingdon, Oxon: Routledge.

Dennett, D. 1987. *The Intentional Stance*. Cambridge: MIT Press.

Dennett, D. 2010. 'The Evolution of Why'. In: B Weiss & J Wanderer, eds. 2010 *Reading Brandom: on Making it Explicit*. Abingdon: Routledge.

Macbeth, D. 2010. 'Inference, Meaning, and Truth in Brandom, Sellars, and Frege'. In: B Weiss & J Wanderer, eds. 2010 *Reading Brandom: on Making it Explicit*. Abingdon: Routledge.

Rosen, G. 1997. Who Makes the Rules Around Here? *Philosophy and Phenomenological Research*. 58 (1).

Searle, J. 1983. *Intentionality, an Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Weiss, B., and Wanderer, J. 2010. 'Introduction'. *Reading Brandom: on Making it Explicit*. Abingdon: Routledge.