Apology and Sincerity

Oskar Lindblom
Magisteravhandling i filosofi
Handledare: Martin Gustafsson
Fakulteten för humaniora, psykologi och teologi
2020

Oskar Lindblom
FHPT vid Åbo Akademi

Table of contents

Oskar Lindblom
FHPT vid Åbo Akademi

APOLOGY AND SINCERITY

BY OSKAR LINDBLOM

When studying the apology as a speech act, the notion of sincerity becomes central for the investigation. In the literature, sincerity has been understood as a matter of correspondence between words and mind. However, the philosophical problems that arise from this definition of sincerity, especially problems belonging to the philosophy of mind, have not been properly addressed. By properly resituating the apology in its ordinary ethical context – where notions such as bonds of trust, undertaking commitments and having integrity are vital – it becomes possible to shift sincerity as a correspondence between word and mind towards sincerity as a correspondence between word and action. Sincerity is thereby anchored in the way speaking entails undertaking certain commitments (and how we can find ourselves going back on our word). Four such commitments that are undertaken in an apology are presented, along with four corresponding ways of going back on one's word: upon saying one is sorry, one undertakes certain commitments regarding justice, agency, normativity and intention.

METHODS AND SCOPE

1 Introduction
1.1 Apology as a question for philosophy

I trace my interest in apologies to bitter arguments that I have had with close ones. The precise details of these arguments have been lost to time, but I can recall the emotional landscapes attached to them: feeling humiliated upon being publicly demanded an apology, feeling wronged upon accusations of carelessness and insensitivity. Those hasty accusations were met with what I now recognize as having been hasty excuses, none of which appeased my interlocutor or served to deescalate the situation. After all, my apologies were not particularly sincere; whether I assumed responsibility for the actions I apologized for was unclear and neither did I identify any particular reason for remorse. Yet I said I was sorry. A new argument breaks out: did I really apologize or not when I obviously did not "mean it"? But I had apologized, I retorted. I said I was sorry and was happy to do so as many times as was required. Counter-retorts were made that the words were there, but an apology was not. What started out as an argument about some small wrongdoing had turned into an argument about what was an apology in my eyes and merely an excuse in theirs. Concepts such as responsibility, remorse and sincerity became subject to heated argument. We knew how to use such words, but what did we mean by them? At the time I had only a vague idea of what a "full and sincere apology" entailed. It was here that I had decided to play an

underhanded trick, saying that if they so please, they are free to lay out the criteria for sincerity by which they get to prove my insincerity. After all, if they judged my utterance insincere, by what measure did they do so? What access did they have to my sincerity? Could they point to it in order to show me that, indeed, the sincerity is not there? Failing this, they would have to take me for my word that I am being sincere in my apology. It was a take-it-or-leave-it offer, turning myself into the one who had been wronged if they chose to pursue the matter further. Rather than admit fault, I had invited skepticism as a trump card and had left it to my interlocutors to stave off if they wanted to make claims about my sincerity.

Looking back, this was a deeply caustic course of action, nothing short of an abuse of language that was made available to me through an acquaintance with philosophy. Only upon closer inspection did I come to recognize where I had originally gone wrong and their initial accusations were right. I expressed my regret, but my earlier challenge haunted me now instead of them: how can I express my regret without leaving room for doubt? I knew when and where to use the words "sorry" and "regret" and while they were not out of place here, how is it that the other can come to know whether I truly am sorry? I figured that if I could articulate the meaning of the words 'apologize', 'sorry', 'sincere' and 'mean it', then I could also *mean them* without leaving room for doubt. Yet upon reflection, I found myself unable to go beyond such doubt. I said I was sorry, knowing I was ignorant to what made those words so powerful. Luckily, my interlocutor did not pursue the matter further. I had said I was sorry and that was that. As my actions lined up with my words, our bond healed over time and our friendship was renewed. But I was none the wiser about apologies and sincerity, at least philosophically speaking.

At times I find myself wondering whether I was *really* being sincere even then. How did they know? How would I know? After all, at least if a gut feeling was to be trusted, I experienced myself as being sincere in the very first apology that was rejected as insincere. But I was not accused of lying about my feelings, but rather accused of lying about what light I viewed the matter in by saying I was sorry. But the problem kept circling back: how did my interlocutor know what I was thinking in order to point out a discrepancy between my words and my mind? What access do they have? How does one apologize sincerely? How ought one apologize? When? To whom? How do I mean what I say? How do I know? How do others know? What does it even mean to mean what one says? On what kind of access can others reprimand me for not being sincere or for not meaning what I say? What exactly is the difference between saying one is sorry and meaning it, as opposed to saying the same words and not mean it? What is it that one really accomplishes upon apologizing? When a speaker emits the words 'I'm sorry', how can such miraculous things happen as 'the speaker apologizes', 'the speaker expresses regret' or 'the speaker takes responsibility'? Can an apology fail, and if so, how and on what grounds? What is it for an apology to be true? Or false? Or fake? Or botched? Or, most importantly, sincere?

We make claims about minds every day, but when everyday claims about sincerity are put on the philosophical chopping block of why we are justified in making them, how are we supposed to defend them? To reject the givenness of such claims and instead demand justification for making them may at first seem like such a philosophical chopping block, but do we not reserve the right to pose this question in the midst of our everyday practices? Would that question simply be out of touch, ill-fated or infelicitous, and if so, why? Or if to seriously invoke skepticism was to shatter the possibility of a practice as a whole, what is it that we have managed to break?

The question that motivates this thesis is the one that crystallizes the experiences above: what, exactly, does it mean to apologize sincerely? To study apologies empirically, separate from philosophical considerations of sincerity, is nothing new: research in linguistic pragmatics has covered some empirical ground on the matter (e.g. Blum-Kulka & Olshtain 1984; Holmes 1990; Lasinkaité 2013; Murphy 2014). But none of these authors reflect on the concept of sincerity beyond acknowledging the importance that sincerity, whatever that may be, might have for the apology. Although such empirical investigation can underscore the importance of sincerity, it does so from outside of the practice that it seeks to describe. It does not *explain* the way in which sincerity is on the one hand understood and on the other hand important for those involved in the practice of apologizing. One could go so far as to say that such investigations into the superficial features of apologies do not bear any necessary link to the way in which we as ordinary language speakers understand what it means to apologize and mean it. The relationship between apology and sincerity is so integral that this thesis cannot follow the neat structure of the two-lane investigation where apologies and sincerity are first defined separately and then related to each other.

But before embarking on answering these questions, the following question must be posed: why should a philosopher talk about something so mundane as apologies? Is it even safe to assume that philosophical questions about something ordinary should make sense to a philosopher? The utterances that make up apologies are not even properly true or false. The positivist worry is that as one attempts to philosophize about mundane ways of speaking, one discovers that the philosophical puzzles that arise from this endeavor actually belong to the bin of "linguistic confusions".[1] Because the subject matter is the language of the everyday, but our framework of understanding them is philosophical, the worry is that questions such as "how do we know if they *mean* what they say?" and "how do we know they *really* apologized?" are so out of place that they are right to be ignored by the philosopher. After all, since we cannot directly peer into the minds of others, utterances such as "I am sorry" and "I know you don't mean it" – utterances that seem to hinge on such a capacity – are actually nonsense. Yet despite lacking the capacity of peering directly other minds, there are right and wrong ways of using those utterances

---

[1] Or, as the Churchlands would conclude, to entertain ordinary language expressions of ourselves as facts to be pursued only leads to contradiction because the manifest image of ourselves that permeates everyday speech is itself contradictory and disjointed.

and have efficacy in our everyday discourse. And if there are right and wrong ways to go about, then it ought to be possible to pose and answer meaningful questions examining the conditions for its success and failure. It is because they still possess such right and wrong ways of use in everyday discourse that they cannot be disregarded as simple nonsense (Hertzberg 2013: 137-138). Although the conventions which govern such right and wrong ways are, so to speak, arbitrary and subject to change, the meaningfulness of the practice in which those utterances do work (and the conventions at play there) must serve as our point of departure rather than a destination of choice.

The aim is to make sense of the way in which ordinary speakers deploy the concept of sincerity in apologies and to thus answer the question "what does it mean for an apology to be sincere?" What, then, will be our source of knowledge? Rather than "count noses" on how people use language within this practice, that knowledge is something that comes with being a competent speaker. Ordinary language speakers, the philosopher included, are already familiar with how to apologize sincerely: the task at hand is to articulate that knowledge. But the articulation of this knowledge remains elusive. The discursive intricacies surrounding apologies are, without a doubt, embedded deeply in our ordinary life. Perhaps too deeply, for this *know-how* is not easily translated into *know-that*. Our everyday familiarity with using language this way is as much part of our linguistic fluency as knowing its syntax or semantics (Cavell 1976). Much like knowing how to speak with proper grammar does not entail the capacity to articulate its conventions, so too are apologies something we know *how* to do without the ability to articulate *what* we are doing. The analogy to grammar can be pushed even further: how to apologize is so integral to everyday life that we do not think so much about the conventions governing those speech acts as much as we think our thoughts and live our lives through them. We face the obstacle of being too familiar with the subject matter, rendering it transparent to us. Bringing it back to view is a task that requires *attention* to our use of language (and by that, I mean attention to how we live our lives as speaking beings) rather than accruing *knowledge* of it. It is thus hard to imagine an inquiry into the nature of something as ordinary as apologies to be carried out in any other field than philosophy, at the intersection of ethics, philosophy of language and philosophy of mind.

1.2 The two main theoretical influences
1.2.1 Ordinary language philosophy

This thesis relies on two philosophical approaches which, while related, are not exactly in agreement. Neither is there any clear line where one ends and the other begins. But this is not a thesis on their relation. They are tools used for teasing out or "disclosing" what is important to the practice of apologizing.

One of the main theoretical influences that this thesis relies on is ordinary language philosophy (commonly abbreviated as OLP). OLP is concerned with tackling philosophical issues without estranging

Oskar Lindblom
FHPT vid Åbo Akademi

our concepts from how they occur in so called "ordinary language". It does so, for instance, by articulating – or philosophizing with a certain sensitivity towards – the ways in which we are and are not inclined to use language and what we achieve with our words. This is the kind of approach taken by Austin in distinguishing the differences between 'accidentally' and 'by mistake', or when Ryle distinguishes when and where it makes sense to describe an action as 'voluntary' and 'involuntary'. When making statements such as "we ask whether … but we do not ask whether …", the ordinary language philosopher points out which what concepts belong to this context and in what context a particular concept would be at home.

An utterance, insofar it is deemed correct, incorrect, appropriate, out of place, in poor taste, called for, misexecuted, even sincere or insincere, is any of those things only against the background of some convention – whether current, desired or hypothetical – against which it can appear as such.[2] But unlike other conventional practices such as filing a lawsuit or playing hockey, there exists no rulebook or step-by-step manual that governs apologies. Rather, these conventions are tacit. Furthermore, while conventional behavior follows from rules in the case of lawsuits and hockey, the same cannot be said about apologies: the rules may be spelled out on the back of conventional behavior, but rarely does the conventional behavior which constitutes an apology follow from some set of rules. Yet speakers of ordinary language "just know" how an apology ought to look like and know how to point out those who do not conform to the conventions that govern it. But when pressed on justifying their claims about what an apology is and isn't, they cannot point to anything besides their own fluency. At times, they may not even have a *common* convention to point to, as all there is to point to is what that convention *ought* to be or what convention best captures the *spirit* of what is being accomplished in a particular speech act, both derived from what they simply consider their grasp of the language they speak. The philosopher is, to a large extent, in the same boat as the speakers populating his examples when doing ordinary language philosophy: all they can point to is their own fluency. Pointing to the various ways in which an apology can appear, as well as articulating the conventions that must be in place for it to appear as such, is not an empirical matter. But this should not discourage us from carrying out an ordinary language investigation into apologies. After all, if we already can make sense of the right and wrong ways of apologizing in our everyday disputes, ought we not as philosophers be able to spell out these distinctions?

The aim is to treat the philosophical concepts that are deployed in apologies without estranging them from the apology as their context. But apologies are not merely a matter of idle chatter. Apologies are embedded in our everyday lives and give voice to everyday ethical concerns such as trust, integrity,

---

[2] Note that making normative judgements regarding language use is, in this case, not so much a job for the philosopher as they are an integral part of the subject matter that the ordinary language philosopher is interested in. Judging an utterance as correct, incorrect, appropriate and so on carried out in everyday speech itself, and the conventions which seem to govern this process in. ordinary language are for the ordinary language philosopher to articulate.

honesty, recognition and sincerity. It is only in this context that the question of sincerity – "how can I know if they really mean it when they say it?" – is a meaningful question. Launching a philosophical investigation into the concepts of apologies and sincerity as *detached* from those practices leads us to general problems in the philosophy of mind and language that are none of our concern when standing before the sentence "I am sorry" in everyday life. One may opt for a view where ordinary language does not share the same goals as "philosophical language"; truth in saying "truth be told…" or sincerity in "yours sincerely" might in fact have nothing to do with the philosophical accounts of "truth" or "sincerity". Such a "traditional philosopher's approach" seems to treat our philosophical concepts as tainted as soon as the practices in which they occur are taken as a point of departure for our understanding rather than being a derivate of our philosophical abstractions.[3] It is the ease and fluency with which ordinary speakers seem to navigate the labyrinthine complexities that the traditional philosopher has painted before himself that is prone to give him vertigo. In worst cases, he escapes into neat abstractions regarding the true meaning of our words from where he can judge our ordinary uses as nonsense. But the cure for this vertigo is, to borrow a phrase from McDowell, "to give up the idea that philosophical thought, about the sorts of practice in question, should be undertaken at some external standpoint, outside our immersion in our familiar forms of life" (McDowell 1998: 63). What our words mean is to be regarded as being equally fickle as the conventions that guide our practices and the only way to understand them is from within those practices. That our words continue to have the meaning they do (or have meaning at all) and continue to have their efficacy in our practices (or have efficacy at all) does not follow from some logical necessity standing outside those practices, but from us

> "…sharing routes of interest and feeling, modes of response, sense of humor and of significance and of fulfillment, of what is outrageous, of what is similar to what else, what a rebuke, what forgiveness, of when an utterance is an assertion, when an appeal, when an explanation – all the whirl of organism Wittgenstein calls 'forms of life'. Human speech and activity, sanity and community, rest upon nothing more, but nothing less, than this." (Cavell 2002: 52).[4]

---

[3] The term "traditional philosopher" is here borrowed from Avner Baz's defense of ordinary language philosophy in the first chapters of *When Words Are Called For* (2012). The term refers to philosophers who hold a certain attitude towards the meaning of a word as being conceptually prior to – and conceptually separate from – the uses in which those concepts are put to work. A traditional philosopher is then what one with Wittgensteinian humor can call someone who not only allows but insists on letting words go on a holiday when we are to analyze their meaning.

[4] I owe the exposure to the passages by McDowell and Cavell to Gustafsson (2004), although I have put these quotes to use to make a different point than he did. However, this thesis as a whole owes its general approach to Gustafsson's treatment of the Cavellian notion of 'pitch' that was explored in said paper. Much like someone who does not have perfect pitch cannot participate in the practice of naming notes by the ear (and justifying that naming), so too would it be impossible for someone completely alien to the practice of apologizing – say, a narcissist or sociopath, or someone from a culture whose apologies we would not even recognize as such – to participate in the line of reasoning that permeates this thesis. But unlike perfect pitch, we *do* expect of people to know how to apologize and we *do* expect to learn until their last breath: what implications follow from apologizing, when and where one can or should say one is sorry, the commitments one undertakes in saying one is sorry, when one is being sincere or not, and so on. And much of our everyday heartache stems not only from people not knowing how to apologize, but from their *refusal* to do so, or their *refusal* to learn, or their refusal to admit that they *do* know how to apologize. Our immersion safeguards us neither from making mistakes nor getting lost in self-deception. Least of all does it safeguard us from betrayal, as it is that immersion which allows for both our trust and others' abuse of it.

What I mean by saying that the meaning and efficacy of our words follow from something other than logical necessity is the following: while our uses of the concepts 'apology' and 'sincerity' have a certain logic to them, that logic is not grounded in something that someone taking an outside perspective to that practice can follow "all the way down" to its logical necessity. At some point, that person will come across a contingency that can only be counted as justifying a practice by already being immersed in that practice. If one wants to speak of a 'logic' that is at work in apologies, to see that logic requires one to already be fully immersed in that practice as part of everyday life. By describing our practices surrounding apologizing and forgiving by phrases such as 'when one says one is sorry for *p*, one implies an acknowledgement of responsibility', what is being produced is a vignette depicting of a form of life, not logical necessity. But forms of life are subject to change. We could imagine a world where, for example, one does not have the right to infer a person's acknowledgement of responsibility from them saying they are sorry for *p*. The fact that such inference follows from saying one is sorry is a matter of convention: that this inference among others hold is the form of life we live in and is perhaps one on which that form of life rests. But while the 'logic' of apologies is in this sense contingent, it is no accidental feature that others draw this inference from someone saying they are sorry: it is against the assumption that such inferences are drawn from what is said that saying one is sorry means what it means. It is only from a point of view that is external to the practice the philosopher describes that the concepts put to use in that practice slip through his fingers as he seeks *justification* for *why* saying one is sorry counts as admitting responsibility.

To make those words not slip through our fingers, we must take their place in everyday language as our point of departure. To first disambiguate the meaning of those words in an abstract milieu where those words do not do any particular work and *then* derive practical applications for it is to put the cart before the horse (Hertzberg 2008). Thus, when something is critiqued in this thesis, it tends to be top-down schematized approaches which, instead of taking the practices in which our concepts do work as a point of departure, seeks to make those practices fit the bill. What usually remains after such approaches is an estranged account of what we do in which we do no longer recognize ourselves doing it.

1.2.2 Speech act theory

On the other hand, this thesis relies heavily on speech act theory (abbreviated here as SAT). SAT is concerned with the nature of actions that are carried out by the means of language. After all, there is a plethora of actions which we carry out by speaking. To be more precise, SAT is concerned with asking what action an utterance *counts as* and the conditions that make it count as such. Take for example the utterance "trust me, the procedure is perfectly safe". Upon uttering, one is on the one hand saying something about the safety of a procedure. One is putting forth a proposition, the truth of which is the

same no matter who says it. But in any ordinary language scenario where one would make that utterance, one would not merely be making a true or false statement about a procedure. Rather, when uttered in a certain context, it counts as *assuring* someone about the safety of the procedure. Whether it succeeds in this endeavor is not logically dependent on the truth of the proposition that the procedure is safe. The interplay between utterance and context is what produces the committing of a speech act. The exploration of how different utterances fair against different contexts is one of the interests pertaining to speech act theory. The ontology of actions can easily turn into a philosophical quagmire, yet in our everyday social interaction, we assume such ontological judgements without second thought: there is little doubt, for instance, that a friend *actually* promised to do something when he said "I promise to do it". Yet for his words to have counted as a promise, there must be some criteria by which we deemed his words to count as a promise. If it did not meet those criteria, it would be in some way botched. Those criteria are known as the *conditions of satisfaction*. It is against the backdrop of these conditions that an utterance can come to count as a specific speech act.[5]

The seminal works on speech acts have been produced by the philosophers J.L. Austin (1962) and John R. Searle (1969), although speech acts have since then enjoyed extensive research in the field of linguistic pragmatics as well. Although Austin provided some initial tools for the study of speech acts, the general framework for such analyses were first developed by Searle in *Speech Acts* (1969). The general framework was later formalized in his joint work with Daniel Vanderveken that produced *Foundations of Illocutionary Logic* (1985), which established the systematic study into the logical foundations of speech act theory. It is here that Searle and Vanderveken settled on a fivefold distinction of speech acts (1985: 179-216): assertives (such as claims and predictions), commissives (such as promises and vows), directives (such as orders and requests), expressives (such as thanks and condolences) and declaratives (such as weddings and adjournments). Since Searle's (1969: 66-67) initial framework for the study of speech acts, apologies have in the literature been unequivocally considered as a part of expressives speech acts, along with verbs such as console, congratulate, lament, praise, greet and welcome (Searle 1969: 21, 30; Searle & Vanderveken 1985: 211-216), despite the fact that apologies *in particular* have received meager attention within the confines of speech act theory. The pervading mode of appearance has been as part of some

---

[5] For the sake of exposition, it is worth adding that speech acts themselves have three components: the locutionary act, which is *to* say something with a propositional content; an illocutionary act, which is the act committed *in* saying something; and a perlocutionary act, which is the effect one achieves *by* saying something. A propositional content, on the other hand, is a technical term that refers to whatever the speech act is about. For instance, the propositional content of "trust me, the procedure is safe" is the safety of the procedure, which is separable from the speech act committed when speaking about the safety of the procedure. Thus the utterances "promise me that the procedure is safe" (when the patient demands assurance), "I doubt that the procedure is safe" (when a colleague reviews the procedure) and "I deem this procedure unsafe" (when the ethical board declares the procedure unsafe) all share the propositional content despite being mobilized as part of different speech acts.

Oskar Lindblom
FHPT vid Åbo Akademi

larger schema of speech acts, such as being included in a list of examples such the one above or as part of an analysis of expressive speech acts (such as Guiraud et al. 2012 and Norrick 1972).

In contrast to my reliance on OLP which is more general, my reliance on SAT is rather explicit. But SAT only serves to denote a starting point, while OLP is the mode of thought that runs throughout this thesis. Apologies, as they fit into SAT, do not give a satisfying answer because they do not actually illuminate the kind of work that apologies and our references to sincerity do in ordinary language. Neither does it illuminate the way in which we *know* whether someone is being sincere or not despite finding a neat place for it among the apology's conditions of satisfaction. To what extent can we claim to be wiser after concluding that apology expresses the speaker being sorry, when we are none the wiser *how* it can come to do so and the ways in which it can *fail* to do so? How do apologies work in ordinary language? What about appeals to sincerity? What does it mean to be sorry? How do we know if someone really is sorry? How do we know if we ourselves are sorry? Rather than counting as philosophical work proper, analyzing the place apologies hold in SAT only serves as the starting point from where can come to do actual philosophical work.

1.2.3 The task at hand

This thesis is concerned with a particular speech act, namely apologizing. But my interest lies not in taking a readymade speech act scheme, such as Searle & Vanderveken's (1985) or Guiarud et al. (2012), as a lens through which I uncritically interpret the speech act in question. Neither am I setting out to discuss the merits of OLP versus SAT as a method nor spend time on their points of contention.[6] They are merely tools for framing the apology as something to philosophize about. At times one needs structure, which the literature on speech acts provides plenty of. At others one has to do philosophy in the way Cavell characterized as treading on the back of having *pitch* for how it is that we speak (Gustafsson 2004). Such an approach may be reproached as eclectic, but the aim is ultimately to philosophize, boldly and emphatically, about what it means to *mean* what one says when apologizing.

By using some of the tools provided by Searlean SAT, I begin by constructing a succinct analysis of the apology in chapter 2.1.[7] This analysis is provisionary at best, as its purpose is to explain the apology's conditions of satisfaction in SAT and serve as starting points for our investigation. One condition in particular appears as bearing significant philosophical baggage. That condition is the so-called *sincerity condition*, which entails that an apology is successful only if the utterance is sincere. In speech

---

[6] For such a discussion, see for example Baz's critique of Searle (2012) and the discussion on Cavell v Searle as inheritors of Austin in Tomas-Fogiel (2011: 43-49).

[7] Although Searle's work is chronologically subsequent to Austin's and greatly influenced by it, I will first introduce Searle's work. The reason for this anachronism is pedagogic: the issue of sincerity is more obvious when imposing Searle's framework (which is more formalized and precise) on apologies than when applying Austin's (which is more rudimentary and ambiguous). However, the notion of sincerity is easier to rearticulate when going back to Austin's work. Like most problems that plague philosophers, the cure is best administered closer to the root rather than its symptoms.

act theory, sincerity is a matter of *correspondence* between the mental state that is stated or implied in the speech act and the actual mental state that the speaker has. Within the context of SAT, apologies, just like most other speech acts, call for a specific mental state in order to be sincere. Whether that mental state is to be called 'being sorry', 'feeling regret' or 'experiencing sorrow', the same question will arise: how do we know when this condition has been satisfied? How do we know that people think or feel as they say they do? The theoretical point that sincerity condition requires a correspondence between words and mental states does not in itself explain how this correspondence is verified when speech acts are deemed sincere. Yet if such a condition does underlie our participation in the practice of apologizing, then the unhindered way we go about within that practice seems to indicate that we are able to verify it with ease. The ghost of skepticism, manifest in questions such as "*how do we know* if someone really is sorry?" and "when are such claims are justified?", haunts the philosopher yet rarely seems to affect those already immersed in the practices where "I am sorry" and its rebuttal do work. We must be speaking of minds on the basis of something, but on the basis of *what* do we do so? How do we explain this seemingly amazing feat? The nature of this correspondence between word and mind remains unclear. Searle skirts around this issue in various ways while Austin's formulation is vague enough to circumvent the issue.

The reason the philosopher has trouble explaining ordinary language speakers' capacity to deem each other's words insincere is because the Searlean account has a certain conception of sincerity – and philosophy of mind more broadly – that misconstrues what it is ordinary language speakers are looking for when judging a speech act insincere. But Searle is not to fault for this. The issue of sincerity is part of what Searle inherits from Austin. Hence chapter 2.2 and onward are characterized by a return to Austin as the site where sincerity can be properly reimagined as another kind of correspondence. The general outline of Austin's conditions of satisfaction is introduced in chapter 2.2 and the apology is analyzed through the lens of this framework in chapters 2.2–2.3. In chapters 2.4 and onwards, the sincerity is reconstructed as a matter of correspondence between words and actions instead of words and mind.

Two shifts of perspective are performed in the course of those chapters, both of which warrant explicit mention in this introduction. The first shift of perspective is from the philosopher's worries of how we know minds *in general* to the kind of worries that are at home in the practice of apologizing *in particular*. After all, an ordinary language speakers' seemingly amazing feat of knowing other minds does not involve solving the general problem of knowing minds that the philosopher burdens himself with. The burden of shifting perspective lies not on the layman to justify his ability to know other minds, but on the philosopher to shift his perspective to make sense of our ability to know other minds. The second shift of perspective is the transformation of our conception of sincerity as a correspondence between words and minds to a correspondence between what we say and what we do. In doing so, to say one is sorry and mean it is transformed from an epistemic question into a much more complex one, involving

moral concepts such as trust, commitment and integrity. What once appeared as an epistemic problem appears now as not really being about obstacles to knowledge at all.

To conclude that sincerity is a matter of correspondence between words and actions does not tell us much. What remains is the arduous task of making sense of this new way of thinking about sincerity in apologies. In order demonstrate how sincerity is a matter of such correspondence in chapters 2.6.1–2.6.4, I first flesh out four concepts that are necessary for this demonstration: the role of trust, being sorry as a cognitive attitude, commitments undertaken in speaking, and going back on one's words (or contradicting them). These concepts are developed throughout chapters 2.4–2.6. B

ut what are the commitments entailed by saying one is sorry? By treating an expressive speech act the same way as one would treat the kinds of sentences featured in Moore's Paradox, one can tease out the implicit features of the apology.[8] By exploring the ways in which we can find ourselves contradicting the utterance "I am sorry" with some other utterance or subsequent action, it becomes possible to both articulate what commitments were implied in saying one is sorry and how they can come to be upheld or broken. In chapters 2.6.1–2.6.4, four necessary elements of a sincere apology are sketched out: injustice, agency, normativity and intention. If successful, the incidental features of the apology are discarded while the essential features remain, leaving us with a portrait of what it means to apologize *and mean it*.

2 Apology and sincerity
2.1 Apology as speech act

Austin's fundamental assumption was that to achieve a proper account of language, an understanding of the link between propositions and truth-value was not enough. Wittgenstein's earlier challenge of that view eventually developed into the Oxford school of which Austin too was a part. To premise one's understanding of language solely on propositions and their truth-value is to fall victim to the *descriptive fallacy*, which means giving undue credence to language that has the purpose of representing states of affairs (Järvenkylä, 2013, s. 78). Sentences that were perfectly functional in everyday speech appear to the philosopher as nonsense when viewed in the light of a this picture of language. But as ordinary language philosophers like to point out, spoken language seems to be composed of utterances with various aims other than just representing the world. What was noteworthy was that many of our utterances which are meaningful lack truth-value. What would it even mean to ask whether "I challenge you to a dance-off!" or "we apologize for the inconvenience" are true statements? If language truly was at its core about making propositions with truth-value, what are we to make of utterances which, instead of being *descriptions* of state of affairs, are the ones that make it so? For instance, what are we to make of utterances

---

[8] Despite contentions with their work in this thesis, this method that is used for teasing out the commitments involved in an expressive speech acts was actually proposed by Searle and Vanderveken (1985: 124-125).

that are used to name, promise, vow, inquire, beseech, bequeath, bet, challenge, fire, make a point, condole, even apologize? These are acts which go beyond describing states of affairs; rather, they bring them about, change them, annul them and so on. The variety of things we can *do* with language was dubbed by Austin *performatives*, primarily explored in the aptly named, and posthumously published, work called *How To Do Things With Words* (1962), based on his William James lectures in the 1950's. Utterances to which questions of truth and falsity could be applied to are the ones that have the purpose of accurately reflecting some state of affairs, which Austin called *constatives*.

Searle, on the other hand, makes the same distinction in another way. Each utterance, insofar one does something by said utterance, is some kind of speech act. Each speech act has a *direction of fit*. Some speech acts – such as stating, reporting, recounting, testifying, and so on – have a word-to-world direction of fit, meaning that if the speech act is to be considered successful, then it is up to the utterance to correctly reflect some state of affair. Others – such as promising, betting, dubbing, firing, and so on – have a world-to-word direction of fit, meaning that if the speech act is to be considered successful, the utterance ought to bring about some state of affairs dictated by the speech act. In order to illustrate this difference, Searle invites us to imagine ourselves being equipped with a shopping list and heading to the store. Meanwhile, someone observing us is taking note of the things we are buying. The list I am using has a world-to-word direction of fit: it is up to states of affairs (in this case my groceries) to match what I intended to buy, i.e. the items on my shopping list. My observer, on the other hand, has made a list with a word-to-world direction of fit. The purpose of his list is to truthfully represent what it is that I shopped at the grocery store. Questions of truth and falsity are applicable to his list, but not mine, because they differ in their direction of fit. The notion of direction of fit seeks to describe which direction causality runs between words and the world.

This notion of direction of fit can be applied to speech acts as well. Some speech acts have the aim of reflecting states of affairs while others have the aim of bringing them about. Determining a speech act's direction of fit is an apt starting point for an investigation into that speech act. When making a statement, the words should fit the world. When making a request, the world is to fit the words. What then is the direction of fit for apologies? When one apologizes, for instance by saying "I'm really sorry I provoked you like that during the seminar yesterday, in front of everybody", is one making an utterance that has the purpose of accurately reflecting some state of affairs, or is the purpose of the utterance to bring something about? Or, to use Austin's terminology, is the apology a constative or a performative speech act?

In order to settle this question, one may be tempted to borrow a technique used by Austin when determining whether an utterance is a performative or a constative speech act. After all, being able to determine if a speech act is constative or performative should also tell us if the speech act in question

has a world-to-word or word-to-world direction of fit. Austin's trick was the following. If one can add the phrase "hereby…" before an utterance without significantly impacting the speech act, then the utterance must be a performative. After all, making the performative aspect of a performative speech act explicit should not impact the fact that the utterance is a performative speech act. For example, one may add "hereby…" in front of "… I promise you that I will be there if you ever need me" without significantly impacting the fact that one has made a promise. But this does not pertain to all utterances. For instance, if one were to preface the utterance "…the forecast for tomorrow's weather is partially cloudy with a chance of rain" with "hereby…", then the utterance takes a sharp turn. What was a perfectly normal forecast has suddenly become something like a decree or proclamation, as if the speaker has the power to decide what the weather will be tomorrow. This comes to show that the utterance in question is not a pure-bred performative, but rather has at least a constative element to it.

Constatives, on the other hand, can be determined by a similar nifty trick. If one were to add "it is true that…" before an utterance without significantly impacting the speech act, then the utterance must be constative. For instance, one may add "it is true that…" before the weather forecast. Even if we do not have the means to determine now what the weather will be tomorrow (it is, after all, merely a forecast), it still makes sense for us to be arguing about what the weather tomorrow will be and go "it is true that the forecast for tomorrow's weather forecast is partially cloudy with a chance of rain". If one prefaces "…I promise that I will be there if you ever need me" with "it is true that…", what would the result be? An utterance that was once a perfectly functional promise has suddenly turned into something one would say if one were to be arguing about whether one has promised or not at some earlier point in time. Prefacing the utterance with "it is true that…" has significantly altered the speech act being committed. After all, a description of one's behaviour and to actually engage in that behaviour are not one and the same.

Austin's trick is so effective because it reveals what direction of fit an utterance possesses. The utterance "I promise that I will be there if you ever need me", which is a commissive pertaining to the speaker, has a world-to-word direction of fit, meaning that it is up to the speaker's actions to line up with what they have committed to in their promise. The utterance "the forecast for tomorrow's weather is partially cloudy with a chance of rain", which is a statement about the weather forecast, has a word-to-world direction of fit, meaning that it is up the speaker to say things that are true. If the same trick is applied to apologies, one ought to be able to determine its direction of fit.

Before moving on to determining the direction of fit, it is worth discussing the reason for doing so. The reason for determining the direction of fit is that it is the most important feature of a speech act along with its *conditions of satisfaction* (the latter often following naturally from the former). The conditions of satisfaction are those conditions which must be satisfied for the utterance to count as a certain speech

act. In *Speech Acts*, Searle sketches four types of conditions which inform us of what is required of an utterance (and circumstance) for it to count as performing a particular speech act: propositional content conditions, preparatory conditions, sincerity conditions and essential conditions (Searle 1969: 63). The propositional content simply restricts what the speech act can be about, for instance what one can promise or apologize for. The preparatory conditions state the context and circumstances that must be in place in order for the utterance to count as that speech act. For example, Searle's preparatory condition for thanking states that the speaker believes they benefit from whatever it is that they are grateful for. The sincerity condition specifies the necessary psychological state for the speech act to be sincere. For example, the speaker ought to believe what they are saying when making a statement and intend to follow through when making a promise. Lastly, the essential condition states what the utterance counts as when felicitous. The point of a request is to make the others grant something, while the point of congratulating is to express joy (or at least recognition) for someone else's fortune.

What, then, are the conditions of satisfaction for the apology? Searle does not present in Speech Acts nor in his joint work with Vanderveken any concrete formulation of its conditions, but other formulations can be found in the literature, such as the one used by Murphy (2014: 39):

1. Propositional content: an act done, or to be done in the future, by the speaker or someone for whom the speaker is a formally recognized representative.
2. Preparatory conditions: speaker believes that the apology recipient, or a contextually relevant third party, believes that the act was an offence against the recipient (or someone whom the recipient represents).
3. Sincerity condition: speaker regrets the act or one of its consequences.[9]
4. Essential condition: utterance counts as an apology.

Since the apology are categorized as an expressive speech act, the sincerity condition and essential condition overlap, meaning that the purpose of the speech act is to express the content specified in the sincerity condition (Searle 1969: 65). If normal input and output conditions obtain (speaker is serious and not apologizing e.g. as part of a line in a play or in quoting someone), saying that one is sorry while fulfilling all of the conditions above counts as apologizing.

2.2.1 The issue of sincerity

But this set of conditions does not tell us which direction of fit apologies have. Does it reflect the world or is the purpose to make the world fit the words? Let us assume that it has a world-to-word direction of fit, which suits it at face value: in saying "I apologize for the atrocities committed by my people", the

---

[9] 'Being sorry', 'to regret' and 'feel remorse', all of which can be argued as being expressed in apologies, are used interchangeably in the literature. Although these concepts deserve proper differentiation, such a differentiation is outside the scope of this thesis. The Searlean sincerity condition is understood as picking out a specific 'psychological state' and the issue of sincerity lies with this picking out as such, not the particular psychological state that it seeks to pick out in the case of apologies.

speaker is not an observer to some state of affairs, but rather an active part in bringing about some change in it. However, if one considers utterances such as "I'm so, so sorry for what I did, I really am", it seems as if it has a word-to-world direction of fit as the speaker attempts to relay the information that they *are* sorry. This claim can then be subjected to the question "is it true?" Whether the apology does its job relies on the truth or falsity of this claim. The world-to-word interpretation can even be explained as a word-to-world in disguise: to say that one apologizes implies that one is sorry, and the illocutionary force of the former relies on the truth of the latter. This seems to be the very reason why "but I did say I was sorry" can be accepted by taking it as sincere (or rejected for the lack thereof). The apology's capacity to change state of affairs (i.e. that one has apologized) is hamstringed by the extent in which it also reflects some state of affairs (i.e. that one is sorry). The former without the latter amounts to what is called a non-apology.

But it is here that we run into a dilemma, one which upon recounting is sure to induce frustration and confusion. Even if the apology as a speech act has neither direction of fit, the sincerity condition has still been defined as the speaker having the mental state implied in the speech act, i.e. a correspondence between word and mind. Insofar the apology is either felicitous or not on the basis of the sincerity condition, the speech act must reflect some state of affairs (which in the case of speech act theory is the speaker's mental state) in order to count as an apology. It is here that we are faced with a problem of understanding how this condition can come to be fulfilled: how do we *know* if the speaker really is sorry? How does the speaker know? Most importantly: in my introductory example, how did my interlocutor know that I was *not* sorry? Such epistemic questions would at first not seem to be appropriate for apologies if they had a world-to-word direction of fit. But because the sincerity condition has a word-to-world direction of fit when understood as a correspondence between word and mind, such epistemic questions seem inevitable.[10]

So far, the analysis has hinged on sincerity being necessary for apologizing, which is why an analysis of sincerity is so crucial. I agree with Murphy who concedes that sincerity seems vital for the execution of an apology, as otherwise one would simply be "going through the motions" (Murphy 2014: 39). But is this necessarily the case? After all, a promise *is* a promise even despite not planning on following through, an invitation *is* an invitation despite not wanting someone to come, and so on. Austin and Searle were adamant in their claims that insincere speech acts still count as those speech acts, albeit

---

[10] The reason why epistemic questions are not appropriate for world-to-word direction of fit is as follows. For instance, we do not ask how the priest *knows* the child's name during its christening because the act of christening a child has the world-to-word direction of fit. It brings the truth of that name about, rather than report on it. However, that question can be asked of someone writing a transcript of the ceremony because that transcript has a word-to-world direction of fit. The christening and the transcript are subjected to different kinds of justifications, one having to explain how it has the power to make it so and the other having to explain how it knows what it is purporting. The "because I said so" and the "because he said so" are different as justifications beyond just being in first and second person.

they may be abusive or deceitful (Austin 1962: 10, Searle 1969: 62, Fotion 2001: 26). To promise and not follow through is reproachable precisely because one has promised despite not intending to follow through. But can that assumption really be extended to apologies? What differentiates an apology from a non-apology is exactly what we perceive as being the sincerity of the utterance. Unlike condolences which are successful even in the absence of compassion and promises which are successful in the absence of intention, apologies do seem to require that the speaker *means it* in order to have properly apologized. Whether or not I could have in my introductory example insisted that I had apologized either way, the epistemic question pertains: how does an interlocutor know if the apology is sincere?

What makes claims about minds so difficult is that they are only superficially comparable to statements about the world. Our statements about ordinary objects and their states of affairs is a fairly straightforward matter, general epistemological contentions aside. While it is possible to check through the window if the weather forecast was correct, one cannot simply peer into someone's mind. Even the highest resolution brain scan is at best beside the point, because claims about minds are not statements about their brains. Whether others speak true of their minds is something we only have their *word* on, yet curiously, our minds are not such a mystery in our everyday conduct.

But on what basis do we deem each other's apologies sincere when we accept them (or insincere as we don't)? For Searle it is obvious: either someone feels as they say, or they don't, much like there either is a goldfinch out in my backyard or there isn't. The conditions of satisfaction, sincerity condition included, are in a perfectly ordinary sense empirical (Searle 1969, p. 178) In the case of claiming there is a goldfinch in one's backyard, one can simply take a gander in the yard to confirm. When deeming an apology sincere, how are we to do so? The closest Searle comes to explaining how we know the minds of others is implied in his institutionalist perspective: there are certain conditions which must be met for an utterance to count as an expression of a certain mental state (e.g. any expressive speech act) and only then can the truth of that mental state be said to be the case. But when the apology is the way for the hearer to know the speaker is sorry, and the speaker actually being sorry is a requirement for the apology to be successful, the dilemma is inevitable. Searle seeks to avoid the epistemic problem of the sincerity condition by stating that in performing an expressive speech act, the speaker neither changes the world nor matches their words to it, but rather "the truth of the expressed proposition is presupposed" (Searle 1976: 12). But this "presupposition" is not only unclear, but at times absent in ordinary language: an apology can easily run into cul-de-sacs where sincerity is doubted, one's grasp of the situation is tested, or where a relationship of trust is not what one thought it was. Neither is the answer to having one's sincerity doubted that "its truth-value is to be presupposed" an appropriate one. To simply skirt the issue

of the epistemic question by suggesting that interlocutors presuppose the truth of "I am sorry" leaves us none the wiser unless the nature of this presupposition is expounded upon.[11]

Rather than conclude that one's interlocutor does not know and *cannot* know such things because of *philosophical* contentions, Searle correctly draws the conclusion that in ordinary language we do speak of minds and justified in doing so. It is only on the back of such a capacity that the sincerity condition can make sense in the grander scheme of Searle's philosophy of language. But the problem lies in the definition of sincerity that is employed in the sincerity condition, namely sincerity as a correspondence between word and mind. Every definition of correspondence poses the question of what access speakers possess when making claims about such a correspondence, such as making the claim that an apology is insincere. This definition of sincerity has been disseminated across all corners of speech act theory but has never been exposed to serious philosophical reconsideration (e.g. Norrick 1976; Searle & Vanderveken 1985, Guiraud et al. 2011). Throughout the rest of this thesis, I seek to shift the definition of sincerity away from a correspondence between word and mind to a correspondence between word and action. Only this way can apologies and sincerity be anchored in their natural habitat: the complex ethical concerns that permeates ordinary language.

2.2 Infelicitous apologies

Before moving on to how insincere apologies can be understood as a correspondence between word and action, it must be pointed out that insincerity is only one type of infelicitous speech act among many. It was in conjunction with the various which in which speech acts can fail, be botched, go awry and so on that Austin formulated speech acts' conditions of satisfaction in *How To Do Things With Words* (1962). It was on this seminal work that Searle came to construct his speech act theory. As we shall see, the issue of sincerity is still present in Austin's framework, but he also leaves "enough tools on the floor" to cobble together a new understanding of the sincerity condition.

Austin proposed six conditions that an utterance must pass for it to count as a "felicitous and non-defective execution" of a speech act. By articulating those conditions, one can analyse the various ways in which illocutionary speech acts can fail or go awry (Austin 1962: 14-15). These served roughly the same function for Austin as conditions of satisfaction came to serve for Searle, but in Austin's schema they are expressed in such a way as to cover *all* speech acts (rather than requiring a unique set of conditions for each speech act).

---

[11] This expounding work is performed in chapter 2.3.2 and onwards in the form of trust and the way words are binding.

A.1: There must exist an accepted conventional procedure having a certain conventional effect, that procedure to include the uttering of certain words by certain persons in certain circumstances, and further,

A.2: the particular persons and circumstances in a given case must be appropriate for the invocation of the particular procedure invoked.

B.1: The procedure must be executed by all participants both correctly and

B.2: completely.

C.1: Where, as often, the procedure is designed for use by persons having certain thoughts or feelings, or for the inauguration of certain consequential conduct on the part of any participant, then a person participating in and so invoking the procedure must in fact have those thoughts or feelings, and the participants must intend so to conduct themselves, and further

C.2: must actually so conduct themselves subsequently.

From these conditions, Austin composed the following taxonomy to illustrate the ways in which illocutionary speech acts can fail.
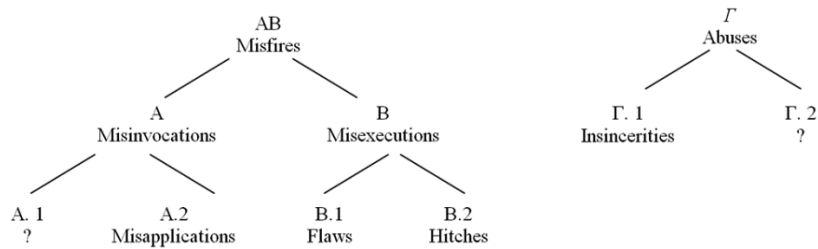

Fig. 1: Austin's taxonomy for ways in which an illocutionary action can fail. (Austin 1962: 18)

This taxonomy shall serve as a "roadmap" for our subsequent analysis of apologies and sincerity. The following task is to go through each of these categories and see how apologies can be viewed through the lens of each infelicity. It is especially the infelicities falling under *Abuses* that interests us because it highlights the issue of sincerity, but the other infelicities are worth discussing first if we seek to understand apologies as whole. This allows us to at least be mindful of which the elements belong to misfires and which the elements of belong to abuses, although the line will often appear blurred. I will be going through the categories in the following order: A.2 (*Misapplications*), B.1 and B.2 (*Flaws* and *Hitches*), after which I will be exploring the kind of infelicity that falls under A.1 (dubbed here *Antinomies*) before moving on the *Abuses* and the way its subgroups are related. I will then argue that C.1 (*Insincerities*) actually cannot be understood without taking into account C.2 (dubbed here *Contradictions*). The reason for doing so is, again, pedagogic: in order to highlight the ways apologies fail in general without losing sight of the issue of sincerity, the infelicities falling under *Misfires* are explored before *Abuses*. Section A.1 is featured after A.2-B.2 because it covers more general ground than A.2-B.2 while also setting the stage for the analysis

of subsequent infelicities. Once the analysis progresses to *Insincerities* and *Contradictions*, the notion of sincerity is explored more broadly, introducing the concepts of 'trust', 'commitments' and 'implication'. The purpose of doing so is, on the one hand, to reconstrue sincerity as a matter of correspondence between words and actions, rather than words and mind. On the other hand, this thesis does seek to go beyond what Austin might have had to say about infelicitous apologies.

2.2.1 Misapplications

A shortcoming in A.2 entails a misapplication. A misapplication occurs when the utterance is applied in a context or with a propositional content that is not appropriate for the speech act one seeks to perform. For instance, a bequeathal can be misapplied in various ways. If one tries to bequeath one's neighbor's house, then the utterance is obviously misapplied if the power to do so belongs solely to one's neighbor. Or, if one tries to bequeath one's child to the neighbor, then the utterance is equally misapplied, since one's child is not an appropriate object for a bequeathal. Nor can one bequeath something once it has already changed hands or bequeath it to the idea of number nine. Although anyone capable of speaking can utter the words, it is only in certain circumstances that those words come to count as a bequeathal with all its judicial consequences. The who, to whom, the what and the when are necessary restrictions that determine when the utterance has been correctly applied and when it has been misapplied. The misapplications are in a sense what was to become the propositional content condition in Searlean speech act theory.

What about misapplied apologies, then? First of all, the propositional content of an apology must be right. What one apologizes *for* should first and foremost be perceived as negatively affecting the person one apologizes. An apology for something that would normally be the cause for celebration or congratulations – or an apology for something that does not affect others at all – is obviously out of place. It has been misapplied. In fact, apologizing for something that is the cause for celebration only serves to imply that the speaker sees something sinister in the propositional content. To bear the burden of one's words having such an implication is the price one must pay to be considered a person capable of speaking. Saying one is sorry for having brought a child to this world, although it may have been misapplied, may lead others to conclude that there must be something wrong with this child or this world. The apology always implies some unhappiness of the matter. And the speaker cannot help this implication: it is simply part of the restrictions placed on the propositional content of a meaningful apology.

But the adverse effects of what one apologizes for is not the only restriction placed on the propositional content, merely the most obvious one. More importantly, the propositional content must

be something one bears some responsibility for but lacks the power to undo.[12] Let us first go through the notion of responsibility before discussing the power to undo, although the two are obviously related. For example, one may utter the words associated with apology when apologizing for being left-handed, but does this amount to a proper apology? One might use words such as "sorry" regarding the inconvenience one's left-handedness causes others, even admit that one's handedness has serious consequences. But is handedness an appropriate thing to apologize for? After all, one would not be inclined to use words such as 'regret' or 'remorse' about the fact that one is left-handed, since one is hardly left-handed by choice. One might be inclined to say "sorry, but I'm left-handed" when a friend goes through the trouble of acquiring a rare guitar that turned out to be right-handed, but in that case the apology is no longer *about* being left-handed: it is about having made one's friend go through the effort in vain. Assuming one did ask a friend to go through this effort, the apology is not about being left-handed but about being careless with one's instructions. That this utterance counts as apologizing for one's poor instructions is not explicitly stated, but rather implicitly understood from the fact that apologizing for being left-handed *does not make sense* in a world where we do not choose our handedness.

But even if the guitar was a gift, acquired and delivered without one's request, the word "sorry" would still be at home. The word alone does not make the utterance an apology when applied outside to a propositional content that lies beyond one's sphere of agency. For instance, both consoling and apologizing share the word "sorry" (as well as the implication of an adverse effect) but they differ in the kind of agency that the speaker assumes. To apologize for something which cannot be placed within one's sphere of agency leads to the apology being misapplied, just like consoling is misapplied when one is also responsible for that injustice. If one says "sorry, but I'm left-handed" when receiving a right-handed guitar as a gift, we know that these words count as condolence, not an apology; again, the reason is that apologizing for handedness does not make sense in our world.

To be *forced* or *expected* to apologize for something outside one's sphere of agency, such as one's handedness or having a certain skin-color, would amount to an insult. One does not choose one's handedness or skin-color. How can one *mean what one says* when apologizing for such things? And what would it be like to be forced to apologize without being able to see oneself meaning it? Although the nature of this tension is unclear, it reveals what has been suggested so far: the propositional content of an apology must be something over which one bears significant agency. Conversely, to apologize for something *implies* that one assumes this point of view on the matter. To be forced to apologize for something what falls within one's sphere of agency is to be forced to extend that sphere to cover it. It is

---

[12] This is a more nuanced formulation of the propositional content condition than the one proposed by e.g. Ogiermann (2009: 46), which simply states that the propositional content condition is a "past act A done by S" without any further specification. As pointed out by Murphy (2014) this does not cover cases where apologies are called for despite the speaker merely being responsible without having performed the action they are apologizing for.

for this reason that we reproach our friends for apologizing when we consider them to be innocent. They should not apologize for crimes they did not commit, nor should they apologize for that which they cannot be faulted for, lest they seek to assume the burden of responsibility.

It is because of this inherent link to agency that apologies focus on the past. If one were to apologize for something one will do in the future, then the reasonable course of action would not be to apologize for it but rather to not do it all. Let us imagine someone who says, with a straight face, "I am sorry, but I have to kill you". What could their words mean? That they are sorry for what they have to do, surely. But if that future truly is inevitable, then would it be appropriate to apologize for it? After all, to say that one is sorry *but* that it is inevitable makes the utterance a defense prefaced with the word 'sorry' rather than an apology followed by context. To say that one is sorry implies some kind of choice, a choice which in apologies is assumed to lie in the past. What one apologizes for must be in the rather specific cross-section of actions which one is responsible for yet do not have the power to undo; actions in the past are chief among them.

Let us move from appropriate propositional content to appropriate context. First of all, the apology must also be uttered by the right person. The one who apologizes should, again, be the one bears some responsibility the matter one apologizes for. However, we sometimes do apologize on the behalf of others. This does not discredit the notion that the responsible should be doing the apologizing. Rather, there is variance in when it is the person responsible who should do the apologizing and when an associate's apology on their behalf is enough. For example, one may apologize for the rude behavior of one's parents and depending on the actions one apologizes for, this can prove to be enough (it can also fail to be enough when the graveness of the infraction is underestimated). If the infraction is grave enough, the person affected might only accept an apology from the person who bears responsibility for the actions in question. The purpose of doing so can be many: perhaps to assure themselves that those responsible understand what they have wrought, or perhaps it is an offer for reconciliation if those responsible have the wisdom to seek it. At other times, the reason for this insistence might be solely vindictive, to strip those responsible from the ethical buttress of other people willing to bear their sin. Nevertheless, the proximity that the speaker must have to the locus of agency seems to scale with the gravity of the situation. One apologizes for their associate when they make minor infractions, but that person's own word is the only thing that counts when the infraction is grave.[13] Searle and Vanderveken

---

[13] But there are also more complicated cases where the interaction between who apologizes and what is apologized for is not so straight forward. For example, the role of apologizer cannot be delegated among a group of friends if the issue is grave enough: when a group of college students are accused of gang rape, wouldn't it be out of place for one of them to step forward and assume the burden for their actions as a group? In this case, he does not speak for the group. Yet when corporations make infractions, this task is regularly delegated upwards and concentrated on fewer individuals the graver the issue is. When the infraction is minor, such as the mislabeling of ingredients on a product, the responsibility is dissipated, but when the same company is accused of a catastrophic oil leak, the responsibility runs uphill for the CEO to bear. In this case, the CEO indeed does speak for the group. It seems as if the notions of collective and public apologies (which sometimes, though not always,

leave this possibility open by stating that an action which a speaker 'is otherwise responsible for' is a legitimate source of an apology (Searle & Vanderveken 1986).

There are even limitations as to whom the apology can be directed. Given that apologies are provided for something that one considers an injustice, the apology should be provided to those whom have been the victims of that injustice. After all, if one apologizes for having sexually harassed a woman, that woman is the person one would apologize to; if that same action was admonished for having sullied her family name, then that apology would be directed to the family as a whole.

But this does not mean that every apology *must* be uttered to the person one wishes to apologize to in order for the utterance to count as an apology. After all, sometimes that person is not present, cannot be present, while in others it is unclear who that apology should be directed to. What I want to draw attention to is that although only victims can forgive wrongdoers, an apology can be legitimate in the absence of those reserving that right. In therapy, the apology might be addressed to whomever it is that one seeks forgiveness from, yet it is the therapist who serves as the placeholder for that person. When making an apology in the memory of the deceased, the dead are only there in memory. In a public apology for the actions where both perpetrator and victim belong to the past, none in the crowd is the right person to step forth and grant absolution.[14] Even confessions, both legal or religious, can be sincere and meaningful without the presence of whomever it is that is supposed to forgive them. In the case of a catholic confession, it is not the priest who forgives the one apologizing: that right is God's alone. In a court room, if the murderer answering to the victim's mother came to voice an apology to the victim as if her memory still lingered, then this apology would not be perceived by the jury as being out of place. In fact, such an apology shows the jury what it is they were looking for: the guilt of having robbed a person of their life, rather than the guilt of having robbed a mother of her child. While the mother can come to forgive the murderer for the latter, it is up to none of them to forgive him for the former. The purpose of the therapist, priest, memorial, court and so on is to facilitate the conditions where a sincere and meaningful apology can be delivered as a form of *catharsis*. Although apologies are revered for their potential to heal and reconcile relationships, it is ultimately the opportunity for a *change of heart* that should be emphasized here. It is in these cathartic cases that the interaction between the speaker and her

---

go hand in hand) complicate the matter, but unfortunately a full picture of their importance are outside the scope of this thesis. For one such discussion, see Murphy's speech act analysis of public apologies by politicians (Murphy 2014).

[14] Apologies by groups of people to other groups for past injustices deserve special mention. For example, the public acknowledgement of the issue via public apologies has been presented as a part of slavery reparations (such as David Pott apologizing for slavery to the descendants of African-American slaves in Maryland, 2004). Such apologies have also been considered as not constituting proper apologies, not because they lack the right emotional valence but because the people apologizing and the ones apologized to do not stand in the right relation to what is being apologized for (see Schedler 2007). Is this not a case of misapplication? It might fulfill a reconciliatory role, but those apologizing are not the ones responsible for what they are apologizing for, and those addressed are the secondary victims of what is being apologized for. The fact that this apology is considered "successful" in achieving this goal does not mean that it was correctly applied: the same could arguably be achieved by expressing one's support for those caught in the aftermath of slavery, committing to undoing its systemic remnants and swearing on fighting such injustices cases in the future.

circumstance defines whether an apology has taken place or not. It is her undertaking of a point of view – one which paints for her a new path as clearly as what follows from "1, 2, 3…" – that makes the apology hit its mark in these cathartic cases. The difference is that in those cases, those who reserve the right to put that guilt to rest are no longer there to do so.

2.2.2 Misexecutions

According to B.1 and B.2 the procedure must be executed by all participants both correctly and completely. A flaw in the execution of a speech act implies that it was not performed in the right manner, even though other circumstances might have been right. For example, a person requesting something cannot just silently wish for his object of desire. He would indeed be desiring, but would he be requesting? Perhaps he would grunt in the general direction of an object and expect that to be recognized as a request: still, there are certain ways of speaking and acting which must be adhered to in order for one's utterance to count as a request. One must articulate one's desire and do so in a manner that others can recognize as a request, perhaps by prefacing the desire with "may I have" or following it with a "please".[15] There are circumstances where this indeed is enough to be recognized as a request, for instance when lying in a hospital bed while being cared for. The caregiver attends to one's grunts and the movement of one's eyes, interpreting their message to fulfill the patient's requests. But in such cases the grunts and eye movements come to count as a request because it is *in that particular setting* a convention for requesting something. Failing to either act inside the constraints of those conventions or failing to act within those conventions to the end entails a misexecution of the speech act.

An apology is likewise governed by some set of conventions that allow us to recognize an utterance as such. The restrictions put upon what we can apologize for, who can do the apologizing and who we can apologize to are part of these conventions. But there are also conventions which govern *how* one ought to apologize. These conventions do of course vary from culture to culture, situation to situation, perhaps even vary between every two people. Such conventions are constantly under negotiation and misunderstandings concerning them are inevitable.

Is bowing and crying conventions for apologies? Although they often accompany a sincere apology and one can be forgiven upon those two alone, has one truly apologized for anything? After all, apologies are not just a matter of expressing emotions. The combination of gestures seems to lack two things which are fundamental to the execution of an apology: a propositional content which denotes *what* one apologizes for and an illocutionary force-indicating device which indicates *that* the utterance is an

---

[15] Such words are what are known in the literature as illocutionary force-indicating devices, meaning that they serve as something for the hearer to "latch on to" in order to judge what the utterance should count as (Searle 1969; Searle & Vanderveken 1985). These do not necessarily have to be choice of words, as they can also be intonations, melody, gestures, grimaces, pauses, word order and so on. What philosophers tend to forget is that the language which we are first introduced to is one of tone: the kind where the voice often does more work than our choice of words.

apology. For instance, we commonly say "I am sorry for what I said to you yesterday" which contains both, the propositional content being that which one said yesterday and the fact that one is sorry about it being the illocutionary force-indicating device. What entails an appropriate propositional content varies from culture to culture as does the illocutionary force-indicating device. For instance, there can be variety in the amount of specificity expected when articulating what one is sorry for; are apologies meant to be vague and hyperextending the speaker's responsibility or ought one specify minutely what one is apologizing for? There is also variation in the extent one is obligated to offer reparation for what one apologizes for. There can even be variation in whether one is to bow or perform some other physical gesture, or if a verbal apology suffices. And all these variables are subjected to culturally conditioned asymmetry: one might be obligated to bow to another when apologizing, while that person might subject to some very other standards. What counts as proper propositional content and illocutionary force-indicating device varies from culture to culture and situation to situation, but what does not seem to vary, if we are to recognize the speech act as an apology, is that there are some conventions regarding both.

What, then, are the cultural conventions that govern proper execution? That is not for me to answer. Unlike the conventions for properly filing a lawsuit or executing a will, there is no rulebook for apologies. This is not to say that there are no rules. After all, there are right and wrong moves to be taken, right and wrong ways to go about. It is only to say that the justification for why someone ought to act this way instead of that way comes quickly to a halt when one has nothing more to rely on than one's notions of what apologizing means. Yet, the fact that this series of justifications does not end up in a set of explicit rules to which we can point does not mean it only ends in our isolated perspectives. When this series of justifications comes to an end, it does not denote linguistic conventions nor subjective accounts, but to what Cavell in *Claim of Reason* (1991) emphasizes as our common practices, the ways in which we share the same worries, habits and wishes that allow our lives to converge (or at times, come into conflict). Only in our fantasies – no doubt tinged by both spitefulness and bureaucratic glee – do such arguments come to a conclusion via the pointing to a pre-established rule with the efficacy of which we do not need to justify. Determining the conventions on which we rely when determining if an utterance counts as the proper execution of an apology is not possible here. All that can be said here is that there are distinctions to be made in the kinds of conventions at play: on the one hand convention regarding what counts as a proper propositional content, and on the other hand conventions regarding what counts as an illocutionary force-indicating device for the speech act in question. What ticks each of these is best found in our forms of life, not some set of rules that float abstractly from it.

How long can the list of conventions be? In other words, what does it mean for a speech act to not just be executed correctly, but completely? The amount of conventions governing when a speech act has been completed must have an upper limit, but there is no upper limit per se. One can always imagine

a cultural context where one must perform one more gesture, say something more, perform some rite or some such to have properly and completely apologized. Some cultures have very elaborate conventions while others are more streamlined or minimal. Thus the upper limit of hoops that must be jumped through is, technically speaking, infinite. But the lower limit can be articulated: an apology is fully executed when it contains a propositional content and an illocutionary force indicating device that indicates that the utterance is to count as an apology. What those conventions are is surely not the same for me as they are for you.

2.2.3 Antinomies

According to A.1, some accepted procedure and effect that concern the speech act in question need to be in place. For instance, one cannot properly promise to a child who does not understand what a promise is or fails to grasp how they work. Or, a child may banish their parent from their pillow fort, only to have that speech act lack efficacy because the parent does not play along with the child's supposed authority. In short, the conventions governing the speech act – how it is to be performed and when it is called for – must be *recognized* by both parties so as to avoid legitimizing speech acts where only one party engages in the procedure. The conventions in question must be *common*.

Austin considers insufficiencies of this type when entertaining cases where one party dismisses a practice altogether or refusing to partake in it (Austin 1962: 26-28). But this is an asymmetric difference: one party thinks the conventions for this speech act hold, while the other does not. For example, one party in a marriage may seek to initiate a divorce while the other considers marriage to be indissoluble. Or, when picked on a team, the person picked may grunt "I'm not playing". But as Austin points out, it is not obvious that these cases fall under A.1, as they may simply be *misapplications* of the utterance "I divorce you" or "I pick George". Or, it might be a *hitch,* since the person accepting the divorce or being picked on a team is a necessary part for the completion of the speech act. But Austin does not take into account cases where both do recognize the speech act in question, only to differ on the accepted procedures that guide it. Rather than being a mistake in the speaker's performance, this infelicity lies in the lack of common ground. If both enter the discussion with incompatible judgements on what counts as that specific speech act, then they are *both* speaking past each other.

Austin was himself aware that his theory might prove to be inadequate and made no claims on being conclusive (Austin 1962: 14). Perhaps this lack of investment is why infelicities above were left unnamed. But this type of infelicity is far from unimportant and so it deserves a name for the sake of this thesis. I propose to call infelicities of this kind *antinomies*, i.e. oppositions between two valid but incommensurable judgements about what the presiding conventions are or ought to be.

In the introductory example of this thesis, one party's apology was rejected by other on the grounds of insincerity. The apologizing party could have, at that point in time, replied that the other is not in a position to reject it on those grounds. There, the utterance "I said that I'm sorry; you can take it or leave it" indicated that all the cards are laid out; what more is the speaker supposed to do than say they are sorry? It is unclear whether there is any obligation for the other to take the speaker for their word, for they already consider themselves to *know* that the speaker is being insincere. To know this, they must, as Austin alludes, be operating on the basis of some convention for what counts as a sincere apology.[16] Regardless of what these conventions are, there are more pressing matters afoot here: do we have the right to assume from the start, as Austin seems to do, that all parties involved in this example are operating within the same set of conventions? What is lost in making such an assumption? And if the difference lies between sets of conventions rather than between interpretations derived from some common ground, can we ever make sense of speech that is, so to speak, flung at each other from separate trenches? What are we to make of such conflicts?

To apologize is not only to concede to the facticity of information from which it follows that one ought to apologize, i.e. to concede that one has in fact not stayed in touch. To apologize for something is also a matter of conceding to a frame of reference within which it follows that one ought to apologize for not having stayed in touch. But if one is to truly apologize, there is even a further conceding that ought to be made in order to avoid further antinomy: one must apologize in accordance to conventions by which the other can recognize the utterance as being an apology. To summarize these three points of concession, one concedes to a matter of fact, what follows from that fact and what counts as having done what follows from that fact. One may concede that one did what one did, but it does not follow that one ought to apologize for it. One may even concede that one ought to apologize, but not apologize in a way that can be recognized as an apology. In other words, one cannot concede to what one did and that it calls for an apology only to claim that "I am sorry you feel that way" or "I guess I'm sorry" ought to do it. The two are still disagreeing on what *counts* as a proper apology.

It is not clear what one party ought to change in order to meet the other half-way, nor is there any indication as to whether they should. Both disagree on whether an apology is owed and have certain preconceptions of what counts as a sincere apology. But they do not experience their difference as laying in their personal opinions. These conventions belong to what they think applies *in general*. Much like

---

[16] While it is hard to pin down what those conventions are, they are the same kinds of conventions that stop one from meaning one is sorry when saying that one is not sorry: to say one is sorry, as opposed to negating it, is one of the rudimentary conventions for what counts as an apology, no matter how much one "in one's mind" tries to *mean* one is sorry when saying one is not sorry. This can be read as a variation of the question posed by Wittgenstein at §PAGE in *Philosophical Investigations*: what would it even entail to assert that one *means* it is hot when one says it is cold? In speaking about what makes words mean what they mean, would we be speaking of something internal to the speaker such as their intention or something external to them such as linguistic convention? While this distinction itself is part of the "picture of language" that Wittgenstein wants us to go beyond, to argue for the linguistic convention interpretation is beyond the scope of this thesis.

Oskar Lindblom
FHPT vid Åbo Akademi

appropriateness and etiquette, these conventions are experienced as applying in the general sense of Hegel's *Sittlichkeit* (2001: 136 [§151]). But if one does not share the frame of reference in which it follows that one owes an apology, being expected to do so can be experienced as insulting, confusing, tyrannical, belittling, and even childish. This experience calls for standing one's ground when the other has false information, such as a friend thinking one has not tried to stay in touch when one has, but such an expectation can be experienced as wrongful even when those holding such expectations are properly informed. Being expected to apologize for not having stayed in touch can rub one the wrong way *despite* it being true. The call for an apology may be seen as insulting precisely because one does not see one's actions as anything to be apologized for (or worse, one sees such a demand as breaking convention). Ought one not, then, stand one's ground? Or ought one to apologize despite not sharing the other's point of view? Perhaps it is easier for everyone if one just says one is sorry and be done with it. Budging just enough to say one is sorry does not mean one budges enough to change that point of view: one has said that one is sorry without *seeing* the matter as anything to be sorry about. But does this really amount to an *apology*?

Let us imagine that one did budge and apologize for not having stayed in touch *despite* viewing it as a trivial matter. Could one truly be sincere in one's apology? After all, one did not *see the problem* in the light one's friend did. Even being told what the problem is does not guarantee that one is capable of seeing it in that light. Now, the words are there, but there lies a difference in judgements of convention; conventions regarding what a friendship entails, what friends ought and ought not to do, what violations of this bond call for an apology and so on. Whether one or the other is right is a matter that could only be settled within a common frame of reference. But whose frame of reference is that supposed to be? Their disagreement lies in what that common ground is or ought to be, with neither budging in their view. In saying one is sorry without budging one's point of view, the apology remains insincere. Under such circumstances, when one speaks as if viewing the world in way one does not, it *cannot* be sincere.

At times, differences of this kind are understood by the parties involved.[17] One party might see the situation as calling for an apology, while the other party is reluctant to see the matter in that light. To burst over family dinner "in this family we *apologize* for lying!" would imply such a scenario. But such conventions for what one ought and ought not to apologize for are rarely – save for rare cases – set in stone or written in blood. Thus, such an utterance is not so much a description of current convention, but rather declarations of what those common conventions *ought* to be (or what the speaker *wishes* for

---

[17] But the parties involved are not always aware that such a disagreement is afoot. At times, this infelicity may have a more naïve character, with both parties speak past each other rather than butting heads. Cultural clashes are often of this kind; what is considered a prediction in one culture may come across as a promise in another. Our frames of reference are not something we choose to uphold or discard, rather they are something we are endowed with from the past that has shaped us. To undo them requires arduous work, with most changes occurring outside of what we choose.

them to be). The possible replies to these kinds of utterances are many, but the kind of context in which an utterance such as this has its home is, undoubtedly, an unhappy one. But in such declarations, there is a desire at play: a desire for some common convention, common ground, common point of view, a common language. There is struggle at play as well: a struggle between those who see their lies as justified and those who see lying as undermining the ethical commonality that their familial bond provides. The parties are not so much having a misunderstanding about the familial bond as much as they are butting heads over what it ought to be. The familial bond means something different for them, and as a consequence, there is a difference in when apologies are called for.

Although Austin admits that everyday life is far from unambiguous, he makes the assumption, in a way that is far from unproblematic, that all parties normally share a common language. If there are disagreements between speakers, that disagreement lies in what to derive from their common language, not a discrepancy in their language. This problematic assumption is then echoed in Searle when limiting his speech act theory to instances where "normal input and output conditions obtain", having a common language and understanding each other being among those (1969: 57). Any instance outside of that constraint falls outside what Austin is interested in, with Searle inheriting this interest. After all, they are interested in language as it appears in cases where language adequately does its job. To carry out speech acts as part of everyday speech would be impossible without language operating smoothly to some degree. Neither is this smooth operation – making ourselves understood being one such feature – something extraordinary in the sense that every instance of doing things with words entailed a great overcoming of philosophical obstacles. But the assumption of such a smooth operation paints a narrow picture of language as it features in our lives. It makes a problematic assumption of language and life being somehow separate entities, with conflicts in the latter remaining lucid despite artificially stipulating a uniformity in the former. Conflicts in the latter reveal, or is at least intrinsically bound, to conflicts in the former and vice versa. The assumption of smooth operation – i.e. the exclusion of antinomy from the picture of language in which speech acts have their home – excludes large swaths of ordinary language. The excluded language use is the one we engage in when the nature of commonality is disagreed on, misunderstood, caught in a tug of war, subjected to distortion or manipulation, and so on. To assume a smooth operation of language is to assume a broader commonality in forms of life as already being in place when it is in fact not. But there is nothing to be taken for granted in such commonality. Any instance where forms of life converge is a precarious one.[18] At times, common ground is achieved, and so too will there be some form of convergence in both language and life. At others, that commonality

---

[18] However, such commonality in language and form of life can possess certain robustness, but this robustness does also fool us to see that commonality as something we cannot undermine with our actions. At worst, overemphasizing the robustness of forms of life tempts us to mythologize about some Ur-form of life – a sense of humanity in the broadest sense – that one can always appealed to or defaulted to for commonality. Even that broadest sense of commonality can suffer desecration and alienation.

tears, revealing the abyss over which "human speech and sanity, activity and community" is suspended (Cavell 2002: 52). These tears – the various hiccups, obstacles, challenges, misunderstandings, impairments, disagreements, failures, abuses and other infelicities of language – are just as inherent to everyday speech as its smooth operation. The conflicts which emerge in apologies cannot be properly understood without paying respect to this side of language. When a picture of language excludes the precariousness of its own smooth operation, a great deal of linguistic behaviour which deals with that precariousness is left at the wayside. Infelicities of this kind are important to not just apologies in particular, but to the analysis of speech acts and ordinary language as a whole. And so, if not properly comprehended, Austin's list of infelicities remains incomplete.

As a result of this exclusion of antinomies, the speakers populating Austin's and Searle's examples seem barred from articulating, defending and challenging the "conventions of a speech act" mentioned in condition A.1. This blinds both the people inhabiting their examples and their readers from properly contemplating the responsibility they bear for the articulation and enactment of those conventions. Neither ordinary language nor ethical life are a matter of blindly adhering to given conventions that determine the common. They include the questioning of those conventions, doubling down on them, rearticulating them, misunderstanding them accidentally or wilfully, feigning ignorance of them, bending them one's whims, interpreting them in good or bad faith, in one's favour, or in someone else's favour, and so on. What those conventions are is simply a part of a form of life, but in being a part of it, they too are precarious. Sometimes this precariousness is good: it allows us to dispose those conventions which we deem tyrannical, pedantic or simply beside the point. On the other hand, that precariousness entails great pain: conventions for what counts as good, truthful, sincere or trustworthy is foundational for our lives while in turn having no other foundation than the robustness of the life which we base on them. If our understanding of a speech act can only be explored in ordinary language situations where the precariousness of current conventions is barred, then we can never account for conflicts where there is butting heads over what conventions ought to apply rather than accusing others of foul play (with the former sometimes being a way of doing the latter). It is only against a full picture of language that it becomes meaningful to ask whether an apology is sincere or not, whether it *needs to be* sincere or not. Whether someone's apology can be called sincere does come down to how it fares against the background of conventions, but it is because those conventions are precarious, even fragile, that we come to cast suspicion on someone's sincerity even when they "do everything just right". [19]

---

[19] Norrick, too, makes some association between apologies, commonality and its precariousness in concluding that "[it] is essential to the smooth working of society that there be standard means of admitting responsibility, implicating remorse, and forgiving. Without these we would probably be at one another's throats much of the time. In this sense, acts of apologizing and forgiving are more basic and important to society than such acts as thanking and congratulating, which by comparison are its pleasant byproducts rather than functional principles." (Norrick, 1978: 284)

2.3 Insincerities

According to C.1 there is a mental component which must also be accounted for in order for the speech act to be sincere. The nature of this mental component remains elusive in Austin's work, as it sometimes refers to emotions while at other times refers to thoughts or simply *meaning* what you say. But Austin remained adamant on his stance that insincere speech acts are still successful, albeit unhappy or half-hearted (Austin 1962: 16-17). The speech act in question does not go awry simply because the speaker did not mean what they said, felt like they implied to be feeling or thought what they reported to be thinking. Their word still stands, just like how a promise was a promise despite the lack of intention to follow through. Likewise, a statement is a statement despite not believing it and so is thanks a thanks despite the lack of gratitude. Some of these infelicities are shared between the two. Some belong to the speaker alone.

Just like any other speech act, the apology bears a link to some mental component that must be satisfied in order for the speech act to be sincere. In other words, there is something that one must be thinking or feeling when apologizing for the apology to be sincere. Austin and Searle have not dedicated any significant portion of their work to analyzing the link between sincerity and apologizing, but they did dedicate some to analyzing the link between sincerity and promising. If this link between sincerity and promising can be properly accounted for, then perhaps it will shed some light on the relation between sincerity and apologizing.

Austin's stance on the link between sincerity and promising was that the *intention* to follow through on one's words. It is one's word that one has given, rather than one's intention. But this raises some concerns. Is the act of promising really just a matter of uttering words? Is there not an "inward and spiritual act" that is the true act of promising? After all, we would be inclined to say that one can promise privately, without moving one's lips or putting one's pen to paper. If this inner aspect is not emphasized, we can imagine a world where thanks are given and promises are made, without anyone experiencing gratitude or intending to follow through on their promises. Or worse, a world where apologies are provided, but remorse is never felt. One of the anxieties that can haunt us in everyday life is whether there is something behind those words which give them substance. Is the person being honest? Can I take them for their word? Can I trust them without *knowing* what they think? Austin was fully aware of what the moralist would deem as the "superficiality of mere words":

> For one who says 'promising is not merely a matter of uttering words! It is an inward and spiritual act!' is apt to appear as a solid moralist standing out against a generation of superficial theorizers: we see him as he sees himself, surveying the invisible depths of ethical space, with all the distinction of a specialist in the *sui generis*. Yet he provides Hippolytus with a let-out, the bigamist with an excuse for his 'I do' and the welsher with a defense for his 'I bet'. Accuracy and morality alike are on the side of the plain saying that *our word is our bond.* (Austin, 1962 s. 10)

Austin's train of thought seems to be the following: it is because all we have are each other's' words that it is those words which bear the ethical weight, rather than the contents of our minds. It is this gap between ourselves and the minds of others which allow for the welsher to get away with saying the words 'I bet' by referring to what he *thought* when he uttered them. After all, there is no way for us to prove him wrong: the contents of his mind are not for us to determine. It is precisely because of this gap between ourselves and others that speech acts cannot be determined by some *inner* action that corresponds to the *outer* one, nor by whatever emotion, thought or attitude that may lie behind.

But is this the case for apologies? After all, if we were to reproach someone for having apologized despite seemingly not being sorry for what they did (or not recognizing *what* they had done), we would do so to point out that they have in fact not apologized. We would say that they *seemed* as if they apologized, but as this illusory semblance might indicate, it amounted to a non-apology. At times, to seem as if one is apologizing when on is in fact not may be a worse insult than not having apologized at all, being not merely a matter a disagreement but an attempted abuse of trust. We yearn for the person to mean what they say when they apologize, for to mean it is the point of the apology. "What good can some words do without it?" one may ask. Austin's account, however, is staunchly sincerity insensitive, and for good reason: all we have is each other's word. But wouldn't the sincerity insensitive interpretation only provide the non-apologizer with the kind of let-out that Austin feared we would provide the welsher in requiring some inner, spiritual act? After all, if sincerity does not matter to the speech act, then the person who apologizes without being sorry can simply say that you have their word and nothing else. It seems as if the one seeking to know whether someone really is sorry or not is the one left sorely disappointed.

Searle's sincerity condition requires of the speaker to have the psychological state(s) that they are imply in their speech act. For instance, if one is to promise that one will be on time the next day, then the sincerity condition states that the speaker needs to also hold the intention of doing so. Likewise, if a person makes a claim that they will be busy tomorrow, then the sincerity condition states that the speaker indeed believes that they will be busy the next day. The sincerity condition requires a kind of *correspondence* between what one says and one's mind. Insincerity is thus some form of break, discontinuity or discrepancy between what one says and what one has in mind. If the purpose of the speech act is explicitly to *express* something, then this condition is of special importance, as at that point the sincerity and truth of the utterance coincide.

Most speech acts, although not necessarily all, contain some kind of sincerity condition. As highlighted, a promise must have the intention to follow through, and a statement must be believed by the speaker. But there are plenty of other examples: when advising, one must actually think that one is actually providing useful information. When ordering, one ought to actually wish for the order to be

carried out. When thanking, one should be grateful. Intention, belief, thinking, desiring and gratitude are all what Searle calls psychological states, and what I call cognitive attitudes (explored in depth in chapter 2.6.1). A promise is a promise despite not having the intention of following through, but the promise is at that point *insincere*. Same goes for a statement one does not believe, advice one knows to be bad, or an order that one does not wish to see executed.

The purpose of the apology is, according to Searle, to express a particular mental state, just like congratulating, complaining, protesting, flattering, praising, and welcoming (Searle 1979: 15). This puts it in Searle's category of *expressives*. When the purpose of a speech act is explicitly to express a certain mental state, then the sincerity condition would seem to play a much more important role. The reason for this is that the sincerity condition and essential conditions overlap in expressive speech acts (Searle 1969: 65). The essential condition of Searle's speech act theory contains what Austin coined the illocutionary point of a speech act. If the point of a speech act is to express a certain mental state, then it would seem to follow that for this expression to succeed, one ought to hold that mental state. If one is not grateful but utter the words 'thank you', has one truly, in one's heart of hearts, thanked someone? In the case of thanking, the answer would be yes. Austin was adamant that speech acts are, in general, sincerity insensitive, as does Searle (Fotion, 2001: 26). Just as an insincere promise is still a promise, so too is an insincere welcome still a welcome. One does not need to hold a certain mental state in order to express holding that mental state. So, to say that one is sorry while not being sorry is simply to express a mental state that one does not hold. The expression is still an expression for that mental state, and the other is right to assume one holds that mental state. It may be untruthful, deceptive or in bad faith, but this does not mean that the speech act has necessarily failed (Searle & Vanderveken 1985: 124-125). What has been said has been said, and no matter of rectifying will take those words back.

2.3.1 Speaking about Minds

Still, there remains something problematic in saying that the sincerity condition is satisfied only when there is correspondence between the speech act and the 'psychological state' that the sincerity condition of that speech act specifies. For there to be a correspondence between the two, there ought also to be some way of measuring or verifying if things are as one says. When one makes claims about ordinary objects, such as the claim "there is a goldfinch in the backyard", the way we go about to verifying the correspondence between claim and reality is rather straightforward (general epistemological concerns aside). But a correspondence between what an agent says about his *mind* and what occurs in it is not analogous to the correspondence between what an agent says about the *world* and what occurs in it. At face value, they share the same grammatical structure, since "I am sorry" is quite like "the goldfinch is in my backyard" in the sense that it picks out an object and makes some sort of claim about it. Their

difference lies in what we consider to be good reason to say that the statement is true or false. Whether the bird in my backyard is a goldfinch (or that the goldfinch *is* in my backyard) can be verified, but how do we go about verifying our claims about the mind? And on what foundation do we decide to "go along" when others say something about their minds? Whether speech acts hinge on the sincerity condition or not, for that condition to be met, we must already either be relying on some access to the minds of others or rely some other set of conditions on the basis of which we infer the minds of others.

Austin, too, sought to make sense of similar issues in another one of his papers, titled *Other Minds*, presented at the Aristotelean Society in 1946. In that paper, Austin criticized the way philosophers had gone about, since Descartes, in discussing minds and what it means to know what other people are thinking and feeling. Directly intuiting what others are thinking or feeling is of course a barred option and it would be unclear what such a direct access would even entail. Thus, philosophers had been inclined to argue that there must be some other medium through which knowledge of other minds is conveyed. After all, we commonly make statements about our minds which others recognize as true, and they in turn make statements about their minds which we recognize as true. At times we go even further and make claims about the minds of others without assuming any direct access! They are claims which others might agree or disagree with, see themselves reflected in, come to find themselves, reject out of mulishness or reject as a matter of setting boundaries. None of the philosopher's general epistemological concerns seem to strike us ordinary language speakers as particularly alarming. We make those claims and discourse about them without claiming that we can directly feel what others are feeling (although a certain sense of humanity comes with the reasoning that this option is not always barred). We constantly make judgements about the minds of others without necessarily having much to go on in making an informed decision. In this sense we say that we "know" the minds of both ourselves and others without necessarily experiencing any complex calculations that lead us to that judgement.

A more recent interpretation of Austin's philosophy of mind has highlighted the links between speaking of one's mind and that of giving a *testimony* (McMyler 2012). I, too, will be pursuing this link, insofar likening apologies to commissive speech acts (such as testifying and promising) reveals what kinds of commitments follow from the utterance "I am sorry". In *Other Minds*, Austin presents an analogy between saying "I know" and "I promise", the latter of which McMyler demonstrates as being analogous to testifying. If all we have is someone's word that *p*, then "knowing that *p* based on a speaker's testimony involves *believing the speaker*, where believing the speaker that *p* cannot be construed as a matter of coming to one's own conclusion about *p* (McMyler 2012: 115). I will seek to extend this analysis to cover the utterance "I am sorry", likening this utterance (and the apology it constitutes) to promising and testifying – and hence construing it as an invitation to trust – rather than making the kinds of assertions which are held to such epistemic standards as justification and proof. If successful, one can start to paint a picture

where answering the utterances "I am sorry" and "I know that you don't mean it" with a call for *proofs* appears as baffling as we would experience it in everyday life. I will first consider the way we speak of our own minds before discussing the way we know the minds of others.

Unlike other philosophers who hold that our relation to other minds is "perceptual", Austin distinguished our ordinary relation from minds from the kinds of non-inferential cognitive relations to material objects characteristic of ordinary perception. Whether we believe what we see in my backyard really is a goldfinch is a matter of impressions: do we believe our eyes or not? And have I properly understood what marks a goldfinch as distinct from other birds that might share its color or size? When speaking about the world, our methodology is very different from when we speak about the mind. When at odds, we look out the window together, compare what we see to, say, pictures in a bird watching book or compare it to other birds in the yard to highlight its "goldfinch-like" features. But in matters about the mind, there is no "we" who can, so to speak, stand side by side at a window and peer together into the world of which we speak, pointing things out and compare them with each other. What we say about the mind is subjected to worries that are very different from the worries that may plague us when we speak of the world: I come to doubt the very person with whom I was supposed to look, rather than doubt my senses which allow me to look. When disagreeing about the mind, our mode of justifying, verifying, doubting and finally settling things comes to a halt at a very different place than when disagreeing about the world. And so, too, must that disagreement call for very different courses of action.

As noted earlier, apologies do not reflect the world like claims or reports do despite the grammar of "I am sorry" suggesting otherwise. It is not, as has been understood, a matter of making true or false assertions and to argue otherwise would invite the problems which arise when we call statements true or false without first being clear on how we verify them as such. To say that speaking one's mind is something beyond true or false statements is simply the most straightforward way to stave off those problems. But it does not clarify what speaking about minds, whether our own or others', actually entails. When someone speaks their mind, their word is the limit: no horizon of verification seems to exist beyond it. All speech acts that Searle refers to as assertives have been ruled out, for apologies do not share their word-to-world direction of fit.

Instead, Austin suggests that speaking of one's mind is like making a promise. To say "I know how to do this" may be taken as a description of an object – the I being the object and having some knowledge being its attribute – but does this reflect the way saying "I know" does work in ordinary language? To insist on it would open it to the same worries that have plagued us so far: how do you know? How do I know that you know? How can that statement be true or false when we cannot peer into each other's minds to see if that knowledge is there or not? Rather, it is much more natural to interpret the utterance "I know how to do this" as giving *assurance* that one knows what one is doing. To

assure someone that one knows what one is doing is a kind of promise, a call for the hearer to trust the speaker's competency. It is not only to let the hearer know that they can rely on the speaker, it is an act that invites the speaker to do so; the speaker shall carry the responsibility. If the speaker then was to be held against the standard of giving *proof* for what they say, there cannot really be any other answer than a simple "Watch."

Sometimes, one speaks one's mind simply to express what one is thinking or feeling. This is what McMyler relates to the notion of giving testimony on a matter of fact.[20] While a testimony might seem like it has the purpose of relaying information, this misconstrues what is truly at stake when giving a testimony. It is true that when one gives a testimony, it might leave the hearer wiser on some state of affairs. For example, testifying that one was not at the scene of the crime is something that is considered when determining the culprit. But taking someone's word for it is not the same as them providing evidence that they were not at the scene of the crime. When a speaker gives a testimony, they simply give their word, much like they do when promising. The comparison to promises is of importance here: both promises and testimonies may be accompanied with some justification for why the speaker is to be trusted, but the notion of trust implies some sort of gap in our knowledge, one which can only be bridged by putting some of our epistemic responsibility in the hands of the speaker. Just as with promises, testimonies are invitations to a certain kind of bond with the speaker, one of *trust*. To construe trust as a conclusion one draws from sufficient proof misconstrues the concept as a whole.

If the speaker makes a claim about the goldfinch, they might point to the various aspects of the bird: its colour, size, song, behaviour, if goldfinches are prevalent in this climate in the first place. But when speaking about the mind, there is no such pointing. But could one not point to what Austin refers to as the "symptoms" of psychological states? Being sorry might be accompanied by certain mannerisms that signal the psychological state of being sorry: crying, bowing, trying to repair the damage one has wrought, and so on. However, none of these mannerisms are "proof" for that someone is sorry, as there is no conceptual necessity between being sorry and acting that way. Indeed, someone might draw the conclusion that a person acting that way was sorry. But those symptoms are not on par with that of *saying* that one is sorry. To say that one is sorry is not a symptom *of* being sorry. It belongs to another category altogether. When drawing conclusions about someone being sorry based on the "symptoms" above, the epistemic burden falls on whoever it is that is making that judgement. But in someone saying that they are sorry, they shoulder some of that burden themselves. When the speaker says they are sorry, their word for it is not simply one behavioural evidence among others. The speaker saying that they are sorry

---

[20] While a testimony is much like a promise, it is not entirely the same: one testifies on a state of affairs' facticity, while promising concerns the intention to execute some action. Words such as "I swear" feature in promises and "I promise" features in testimonies, but as with apologies and condolences, the words that feature the utterance do not determine the speech act. There are fine differences in appropriate propositional content, circumstance where the speech act is called for, illocutionary force-indicating devices used and so on, but the finer details of this difference is outside the scope of this thesis.

is not one piece of evidence among others, because *taking them for their word* is inconsistent with that of treating their word as a piece of evidence to be evaluated when coming to one's own conclusion, because one belongs to a second-person account of knowledge and the other belongs to a third-person.

The difference between their own judgements about their state of mind, what they say about their state of mind and the proofs we have for them having that state of mind is the difference between first-, second- and third-person account of knowledge. It is the second of these accounts, the second-person account of knowledge, that has been so neglected in the history of philosophy: the kind of knowledge which we have of others because we choose to trust them. To reduce all knowledge to first- and third-person accounts is to eradicate the sense of communality where speaking of minds make sense in the first place. Rather than our knowledge of each other's minds being sets of third-person account of knowledge, one which each of us are doomed to form in isolation from others, knowledge of other minds is a second-person account: it is one where the speaker invites the hearer to enact a certain bond of trust, a bond which can be heeded, broken, tarnished, soured, respected, deepened and so on.

2.3.2 The role of trust

Let me rearticulate what is being argued here, because my point is admittedly obtuse (and my approach rather oblique). I want to argue against construing our social lives as consisting of atomized subjects, each drawing their own conclusions about the minds of others and being solely responsible for the accuracy of those conclusions. This is in line with Lagerspetz in his argument that trust ought not to be construed in terms of an individual's beliefs and dispositions (1996: 43). Although the minds of others are, at times, something we judge by observing them, this is not the only, perhaps not even the primary, way that we come to know what they think and feel. Most of what we call "knowledge" of other minds comes from what others *tell* us they are thinking and feeling, i.e. their *testimonies*. And although we still do not know if people purporting to think and feel a certain way are telling the truth, we nevertheless gain the confidence to say that we know once those people say what they think and feel. This is knowledge which does not rest on positivist or empiricist conceptions of sense impressions and inferences, but rather it is a kind of knowledge that hinges on us being social creatures. Austin invites us to appreciate this as a "lifeform": one where our personal knowledge and conclusions are a secondary matter that follows from the commitments and relationships that we undertake as part of our social lives (McMyler 2012: 142). If the conclusion that someone is sorry for their actions was simply an inference drawn by the interlocutor, rather than *trusting* the speaker that they see things in a certain light, then apologies would be impossible. The philosophical problems concerning our knowledge of minds are perhaps the product of an unwillingness to admit such dependence on others as a genuine way of knowing. Yet the same problem is motivated by ordinary human concerns about our relationships to other people, the precarious

positions we put ourselves in if we dare to believe them and the ways in which we can be hurt if we do (McMyler 2012: 142-144).

To summarize, what is being argued against here is a certain picture of what it means to know: one where knowing minds is primarily a solitary matter which can lead us to find other people, atomized subjects who are born into this world alone and views the world through the screen of what they personally think, believe and feel. Rather, what I want to argue for is that such a third-person account of other minds is a secondary feature of us humans finding ourselves living in a social world. Rather than consider alienation our point of departure, I would argue that alienation follows the trauma made available by our initial sense of trust. A trust into which we are born and which otherwise, when arduously won over, is a matter of healing and reconciliation. Our worries regarding sincerity – "do they really think and feel as they say they do?" – is grounded in real trauma of having been fooled, betrayed, played for a sucker, or neglected. To construe this as an obstacle for each to overcome before being *justified* in our knowledge of other minds is to put the horse before the cart: such concerns only arise once we have already known those bonds and suffered injury as consequence. As paradoxical as it may sound, feelings of paranoia, alienation, isolation and neglect are consequences of coming to know other minds and the risk of injury that it entails, rather than a consequence of its lack. The alienation is ubiquitous and part of our most ordinary anxieties, but not in the sense that is implied by philosophers who neglect second-person accounts of knowledge (which I would compare to the *methodological individualists* who Lagerspetz (1996) sees himself arguing against in his work on trust). Our ordinary opposite of trust is alienation and paranoia. The philosopher who neglects second-person accounts of knowledge perpetually describes trust in the form of its other kind of opposite: that of constantly dealing with others by devising *theories* as to how they will behave (Lagerspetz 1996: 56). To construe trust as a matter of evidence only amounts to it being a portrait of its very opposite.

For the sake of clarity, it is worth approaching my point from a second angle. I want to argue against the notion that that our trust is grounded in some form of rationality that assesses evidence and formulates reasons for trusting others. In his work on trust, Lagerspetz argues that this assumption misconstrues it to the point where we no longer recognize it as what we do when we trust. While we expect rational persons to seek evidence for the beliefs that they hold and pass that evidence on to others, trust is different. Trust becomes a feature of our language in moments where we do *not* seek evidence, or worse, grow resistant to any evidence that runs counter to one's trust (1996: 44). Such behaviour sports what is classically perceived as *irrationality*, that of being obtuse, biased, mulish, willingly ignorant or self-deceptive, as well as other forms of general cognitive inertia. Yet, as Lagerspetz points out, we not only see such behaviour as natural but desirable. We expect our friends to follow suit when we exercise trust and think highly of those who show exceptional commitment to it. Trust involves, to borrow Lagerspetz'

formulation, *going beyond or against available evidence* (1996: 45). Trust carries a strange relationship to what may be considered evidence for its rationality. One does not trust *because* one has evidence; it is when one takes other for their word *despite* evidence that we are inclined to call their behaviour an exercise of trust (at times a blind or foolish kind).

Furthermore, the whole picture of trust being built on the back of intuitively gathered evidence is nonsensical as evidence is never evidence *as such*. My trust (or lack thereof) already constitutes the frame of reference in which something can emerge as evidence. If I trust someone and take them for their word, evidence pertaining to their sincerity will appear to me in a certain light. If I am not prepared to consider my friend's apology questionable, I cannot treat a fact as a piece of evidence for or against it. "The total picture looks different – and the picture defines 'evidence' for me" (1996: 66). And this is central to those relationships that are defined by trust, such as friendships. The signs that should trigger suspicion in a relationship between strangers *should* not do so here. To let evidence against the trustworthiness of one's friend sway one's trust undermines the friendship in question. Trust is a form of *not* giving in to the sway of evidence; to abstain from what Cavell might have called a form of scepticism.

The difficulty of formulating a positive account of "reasons" for trusting others is, for Austin, not a problem to be solved, but an aspect of life to be coped with. The kind of relationship we enter when believing (or not believing) another human being about how she thinks or feels courts a certain precariousness that is typically not associated with us believing our senses when ascertaining ordinary objects. The invitation to trust puts both the speaker and the hearer in their own precarious position, inviting the kind of risks that only come with that particular bond: being hurt, betrayed, ratted out, ignored, neglected, made fun of, and so on. Although our eyes may deceive us, they do not hurt or betray. But when others deceive us, or lead us on, or opens us up for abuse, that bond can be severed or soured. Yet we cannot live, at least not felicitously, without such bonds. Our willingness (and uneasiness) to put ourselves in that precarious position is a foundational aspect of our lives; it is "non-optional" (McMyler 2012: 142). Knowledge of a person's sincerity comes not from the evidence one accrues in the moment of their speaking *before* judging whether to trust them. Neither is trusting someone as sincere attributable to making predictions of how their words will line up with their actions. They are not taken for their word *because* one trusts them: taking them for their word and forming that bond of trust are one and the same. It is perhaps this that Searle means when he states that the truth value of expressives must be "presupposed" rather than verified: he never elaborates on it in his own work.

Why argue that knowing other minds is based on trust? Because it informs us of what it is we are on the lookout for when asking what can satisfy the sincerity condition. The point is to rule out that which you use your eyes for before choosing to trust someone: their demeanour, tone of voice, choice

of words and so on. Instead, the sincerity must be seen as that of forming a bond which can then be broken, tarnished, soured, heeded, respected and so on. So far, our analysis has been led in circles by the goose chase of *verifying* the correspondence between word and mind and looking for *proof* or *evidence* that could verify that correspondence. In the moment of speaking, the one apologizing has nothing more than their word to back up their word. Our analysis must not be limited to what comes to count as an apology *as it is said*, as if there is nothing more to say once the apology has been uttered. Rather, we must complete the transition of speaking about a correspondence between word and mind to speak about the correspondence between word and action. The apology can turn out to be insincere as time runs its course and words remain just words. Just like a promise does not end at when uttered, so too does an apology only get off the ground at the moment of its uttering. The commitments one undertakes *in* apologizing form the bond between the speaker and the hearer, one that cannot take off the ground without being a bond of trust. But to articulate those commitments is a difficult task.

To speak one's mind commits the speaker to something, and if committed to something, those commitments are liable to the kinds of dangers that any bond between two people is liable to. It would seem as the matter of speaking one's mind spills outside the neat limits of speech situation: one must take into account the way the person speaks and acts over a larger span of time in order to know what to make on an apology. In other words, matters of sincerity cannot be a mere matter of a person feeling as they say they do *at the same time as saying* that they feel that way, because the way we know whether they feel as they say depends on whether they choose to betray the commitments they undertake in saying that they feel a certain way. Apologies are thus not a matter of *mere* expression of what is there in the moment, with no strings attached for the speaker: they are much like promises and testimonies in the sense that they bring us closer to other people because, much like other commitments, they are invitations to trust.

2.4 Going back on one's words

Both testifying and promising are ways in which a speaker gives her word. A speaker giving her word involves her entering a relationship with her audience characterized by the speaker undertaking commitments. A commitment also entails an obligation which restricts the speaker's future set of actions. A testimony, for instance, entails that one does not go back on (or contradict) what they said earlier. A promise involves, quite straightforwardly, that the speaker gives their word for choosing to do something in the future (or choose to omit doing something). The speaker makes something their responsibility: the truth of their testimony, the following through on a promise, that one sees things the way implied in the apology. Commitments are what can lead a speaker to a kind of contradictions down the line: promising

something and not following through, declaring something as free or independent and then disrespect that sovereignty, bequeathing something and then insist on still deciding over it, and so on.

Infractions of this kind, where something is said and then contradicted, belong to C.2 in Austin's taxonomy of infelicities. Austin referred to infelicities of this type with various names, such as "non-fulfilments, disloyalties, infractions, indisciplines [or] breaches" (Austin 1962: 18).[21] The diagram that Austin provided in *How To Do Things With Words* did not include a definitive name for infelicities of this type, meaning the type of infelicities corresponding to an infraction against condition C.2. An infraction of this type could only occur in speech acts which serve to place the speaker under some kind of commitment. That commitment can then be broken further down the line (such as when breaks one's promise), if not further down the sentence (such as when one says "I am sorry *but, in my defense, …*").

However, I argue that most if not all speech acts can fail as a matter of contradiction. Typically, contradictions have been understood as that of a conflict between two irreconcilable propositional contents, such as saying that "it is raining" and "it is not raining". But this only pertains to two explicit propositions that are in conflict. A contradiction can also arise between a propositional content and what is *implied* by the speech act in which that propositional content features, such as the assertion "the ashes are still warm" implying that one also feels the heat of said ashes. Or, the famous sentence that features in Moore's Paradox: "it is raining" implies that the speaker holds the *belief* that it is raining, and thus to follow it up with "…but I don't think that it is raining" creates a type of contradiction between what was *implied* by the first half of the sentence and what was explicitly stated in the second half of the sentence. The contradiction lies not in propositional contents that are expressed in the sentence, but between what is necessarily implied and what is explicitly stated. Another example of this which will be touched upon throughout the rest of this thesis is "I am sorry for what I did *but* I do not regret it". Searle and Vanderveken liken this denial of the psychological state implied in the speech act to Moore's Paradox, i.e. claiming that it is raining but not believing that it is raining (Searle & Vanderveken 1985: 124-125).

But where do commitments enter the picture? Commitments enter the picture when we ask why the contradiction occurs in Moore's Paradox. To say that it is raining implies the belief that it is raining. But the matter of the speaker believing something suffers from the same problem as other psychological states: they seem to be a private affair belonging to who holds the belief. When someone says they know something, believe something, intend something, mean something, and so on, how are we to verify what they say? In ordinary language, we regularly make ourselves understood when we say that we know, believe, intend, mean, and so on. We even make ourselves understood when speaking of *others* doing so. Someone saying, "you are a terrible person" is open to the rebuttal "I know you don't believe that", just as a promise can be rebutted by "I know you have no intention to keep your word". The minds of others

---

[21] Falkenberg (2012) insists on referring to these as disloyalties.

are no mystery to us. Given that our minds – when stated *and* when implied – are commonly doubted, accepted, given a chance, rejected and so on, they cannot be a private matter to the speaker and philosophy does not "reveal them" to be so either. Neither can Moore's Paradox be rebutted by saying that "yes but I didn't mean that I believed that it was raining".[22] Implications for what we do with our minds follow from certain speech acts and others have the right to infer them. Rather than being a matter of psychological state, implication follows from certain speech acts because saying that it is raining places the speaker under certain obligations whether they choose to undertake them or not – the obligation to act in a certain way that is in correspondence with what has been implied. When done knowingly, we consider it the undertaking certain commitments. A speaker saying that "it is raining" has, in the eyes of others, restricted their future set of actions if they wish to not contradict the commitments that they have undertaken by saying that. Even speech acts that did not have the purpose of expressing any particular psychological states still manage to say a lot about our minds via implicature.

What about speech acts which *do* have the purpose of expressing particular psychological states? In Austin's conditions of satisfaction, condition C.1 stipulates that "if the procedure is designed for use by persons having certain thoughts or feelings", then the person invoking the procedure ought to also have the thoughts or feelings implied in that speech act. This is what it means for the speech act to be sincere. It is worth reiterating that, so far, we are none the wiser on how to verify such a correspondence. Condition C.2 stipulates that the speaker ought also to act in accordance with what she implies to be thinking or feeling. It is here that we locate contradictions: in saying one thing and then acting in another.

In any given language, there are certain conventions for what actions count as acting in accordance with (or contradictory to) what one says. When one says one is sad, that language in which that expression is made comes with conventions for which behaviors are in accordance with that expression and which contradict it. Being mopey and distant reinforces the expression, while being jolly and outgoing contradicts it. Or, to say that one thinks a friend's present is lovely can be contradicted by one's subsequent choices: by refusing to display it for others, by chucking it in the bin, by badmouthing their friend for having picked such a poor present. It is when a person says that they think or feel one way and then act in another that we come to either consider the person insincere or that they do not quite understand what their words entail. Just like promising and testifying, to say what one thinks and feels places the person under certain obligations if they wish to be competent speakers. In other words, to say what one thinks and feels counts as the undertaking of certain commitments.

---

[22] Doing so raises the same concerns Wittgenstein had in *Philosophical Investigations* when asking in §510 what "saying that it is cold and meaning that it is warm" would entail (1992: 162). Can one say that is raining and not mean that one believes it to be so?

In order to properly characterize the way contradictions of this type occurs, let us situate it in a speech act that has been thoroughly analyzed in the literature: promises.[23] If one promises to be somewhere the next day, the contradiction occurs not because one says one will be somewhere the next day as a matter of *prediction* and that prediction turning out not to be true. A promise is, after all, not a prediction; one can still promise something which one *knows* to be unlikely to happen and promising something that one *knows* will happen either way is not much a promise. Rather, the contradiction occurs between the implied intention and then not acting in accordance with that intention.

To promise is to give one's word, perhaps even pledge or testify, that one holds a certain intention regarding one's actions in the future. When one acts differently than as said, what proof does one have for ever having held the purported intention? To say that the intention was there fails, because saying that the intention was there is not like saying that the goldfinch really was in my backyard: we can look for its droppings or stay to see if they live in the area, but the same cannot be done for an intention. The truthfulness of such correspondence between word and mind will always remain a mystery. Perhaps one *did* intend to be on time in the moment of promising. But saying so has now been rendered void. To make those words meaningful again (instead of void), apologies are made, and that bond of trust may or may not be reestablished. If I were to promise that I will be somewhere tomorrow and not follow through, others would still reserve the right to accuse me of having promised *without* the intention following through. After all, all they have of my intention is my word, and my subsequent actions are what can render those words null and void. A bond of trust that I established between myself and others has soured, and my attempts at promises might now be plagued by others' skepticism and doubt. As I promised and they took me for my word, what they had was not *knowledge* pertaining to my mind, but *trust* that I understand what I say and what obligations my words put me under. What commitments follow from my words is a matter of *common* convention, not the *speaker's* intention.[24] In promising something and then changing one's mind is a change in intention, but it was not that intention that was the binding element. Rather, one's privilege to say something (and be understood) comes with the

---

[23] Comparing apologies to promises – and perhaps expressives to commissives in general – is apt because just like promises, apologies are a way we partake in *ethical life* via language. Instead of viewing ethics as a matter of normatively right and wrong judgements, the study of those speech acts which make up the ethical fabric of everyday life highlights the way ethical concerns are something that is a tacit part of our everyday conduct rather than being a restriction imposed on it. Apologies, reproaches, promises, declarations, doubts, statements and so on are ways in which things get done when we demand and provide reasons for actions (or omissions). Much like promises, apologies situate the speaker in a complex structure of commitments. But this, I see, is true of all meaningful speech.

[24] In fact, the kinds of misunderstandings that amount to what one person intended with their words and how the other interpreted them is not a matter of the speaker having a better access to their meaning than their interlocutor. They are in conflict about their judgement about the conventions at play, what words mean what things, what is implied by what words, and so on. They are arguing about something which they both have equal access to, which is their common language. When one person says they "should be able to make it" to an event and the other takes it as a promise, the latter is not wrong in their assumption simply because the former did not mean it *that* way. Rather, the latter is wrong because that utterance does not *count* as a promise in their shared ordinary language, regardless of the speaker's intention. In fact, the separation between the common convention and a speaker's intention should be questioned.

obligation to act accordingly; the expectation put upon the speaker to act as they say is what makes what they say meaningful. The person who "cries wolf" will be quick to have that privilege suspended. Learning this exchange of privilege and obligation that comes with partaking in a common lifeworld – those privileges ranging from the ability to express oneself to enacting change in the world, the ability repudiate and challenge others' accounts, to accuse and to atone, to be a reliable informant and so on – involves both learning how to partake in ordinary language as well as partaking in everyday ethical life. To speak is thus not only to express, but also to undertake commitments that allows for bonds between speakers. It is at this junction that ethics is not a matter of theory of right and wrong, but a matter of linguistic cohabiting, exemplified in one's privilege to speak truth being at stake when one betrays the trust of others by lying.

The kind of deception that permeates our everyday life is not that of saying it is raining when it is in fact not. It is rather about leading others on to think one believes something when one does in fact not: to say and not mean what one says, to be half-hearted in one's word, being two-faced, telling white lies, being insincere, and so on. When a person asks if we think their outfit is flattering, they are not angry because we said that they are flattering when they were in fact not. Rather, they are angry because it was implied that we thought they were flattering when we in fact did not think that. The difference here is subtle, but important. The kind of deception that hurts us the most is one that undermines trust: deception that ultimately obfuscates us from each other rather than being obfuscated from the world. And it is precisely when we say we think or feel one way and then act another that the bond of trust becomes strained.

But in saying that words can come to be contradicted by actions, it is worth noting that I do not want to make too much of a distinction between words and actions. The difference becomes increasingly untenable when reminded that words are actions (speaking is itself to *do* something) and actions are words (how we act *tells* others how we see things). To say one thing and then act another is a contradiction in a sense that is similar to acting one way and then denying it in one's words. If one says that one is sorry, one has not simply uttered words: one has testified, promised, even assured that one views one's past actions in a certain light. But to say one views things in a certain light is not a matter that comes without strings attached to it. One must act in accordance with the commitments one undertakes *in* saying that one views things in that certain light. If not, a contradiction between what one has said and done ensues. For example, to commit the same injustices one recently apologized for is not simply a doing: that action *tells* others that one does not view them as the kinds of injustices one implied in the apology.

To act accordingly is more important in some speech acts than others. To act accordingly is of importance when bequeathing an object, because one must also hand over the object one has bequeathed. To say that one bequeaths a family heirloom yet refuses to hand it over (or destroys it or gives it to

someone else) is still a bequeathal; the speech act is intact but has been ruined by subsequent squandering. One *has* bequeathed the object and that is exactly why one is reproached when subsequent actions contradict that bequeathal. Likewise, one *has* promised when giving one's word and that is exactly why one is reproached when one does not follow through.

However, one might say, apologizing is not a speech act like bequeathing and promising which Searle calls *commissive* speech acts, i.e. those speech acts which serve to commit the speaker to certain future actions or dispositions. Apologizing belongs to the same category as thanking, welcoming, complimenting and other *expressive* speech acts, i.e. those speech acts which serve to express the speaker's feelings about themselves or the world (Searle 1976: 12). But in order to understand the sincerity condition involved in these speech acts without relapsing into the old problems of how we *know* sincerity, that commissive nature must be extended to concern expressive speech acts as well. Saying one thinks or feels a certain way puts one under some kind of expectations of how one ought to act; it is through a coherence between everything the speaker says and does that we know if we are to take them seriously when saying how they think and feel. If the speaker commits themselves via expressive statements of feeling one way, only to then act in another, we reserve the right to question whether that first expressive speech act was "true" after all. Or, rather, we reserve the right to accuse them of being insincere or fake, or as not knowing exactly what the words they used mean, or as being abusing or stretching words in their favor or using them willy-nilly.

2.5 What does it mean to be sorry?

However, before moving on to the commitments one undertakes in apologizing and the ways one can contradict an apology, I want to express some contentions with what it is that the sincerity condition *refers to*. The sincerity condition demands as a correspondence between the psychological state expressed and the psychological state of the speaker.[25] In apologies, that psychological state is called 'being sorry'. Since we are keen to point out when this correspondence is lacking, it can be concluded that we, in the course of everyday speech, know what 'being sorry' means. But what does 'being sorry' mean?

One course of action would be to first narrow down the investigation to answer what kinds of psychological states belong to which kinds of speech acts. Following the fivefold distinction of illocutionary speech acts provided by Searle and Vanderveken (1985) of assertives, commissives, directives, expressives and declaratives, Falkenberg (2012) identified assertives as implying *beliefs*, commissives as implying *intentions*, directives as implying *desires* and expressives as implying *emotions,* while declaratives did not have any particular sincerity condition. Following this schema, the apology expresses

---

[25] Psychological states is the term used by Searle, but the terminology is far from being set in stone in the literature. For instance, Austin speaks of what we think and feel, while McMyler refers to it as mental states.

some form of emotion as the content specified in the sincerity condition, given that it is categorized in the literature as an expressive speech act.[26] This is already something of a lead: being sorry, which is the psychological state with which the sincerity condition of the apology is concerned, is an emotion.

In the case of apologies, Searle (1976) suggests that the psychological state expressed by apologies is *regret*. In a more recent approach, Guiraud et al. have reinforced this notion by suggesting that expressive speech acts are public expressions of emotional states (Guiraud et al. 2011: 1031). Guiraud et al. (2011) argue that the emotions corresponding to Searle's 'psychological states' mean emotions only when belonging to an expressive speech act (2011: 1035-1037). In their approach, they outlined the cognitive structures characterizing the emotions which underlie expressive speech acts, going so far as to suggest that all emotions can be expressed in terms of beliefs, goals, ideals and responsibility.[27] The reason for doing so hinges on Searle's assertion in *Speech Acts* that whenever there is a psychological state specified in the sincerity condition, the speech act *counts as* an expression of that psychological state (Searle 1969). Regret, for Guiraud et al., is an emotion, but for them all complex emotions (which includes regret) ought to also be able to express in terms of beliefs, intentions and desires in order for expressive speech acts to be meaningful. Regret in particular refers to the speaker's belief that they have done something which betrays their ideals or desires (2011: 1035-1037). An apology is simply the codified expression of this particular set of beliefs and desires.

What Guiraud et al. get right is that the psychological state expressed in the speech act is more than just an expression for the valence that colors the speaker's the phenomenological experience, i.e. the way 'being sad' colors the world in which one finds oneself. After all, such an expression would typically fail to qualify as an apology: it lacks admittance of responsibility, denial of the outcome's inevitability, a certain account of agency, public recognition of injustice, commitment to avoiding that action in the future, among other things. In other words, the expression of a momentary gut feeling is not enough to count as regret or being sorry, much less an apology.

Is there a difference between regret and being sorry? After all, the literature seems to suggest that the psychological state associated with apologies is regret while this thesis has featured being sorry as its primary psychological state. Throughout this thesis, I will be assuming that being sorry already implies some form of regret but not vice versa. To say that one is sorry for what one did implies something more

---

[26] As implied so far, there is reason to doubt the neatness of this schema. An apology has assertive features insofar it is a statement about minds, commissive insofar it puts the speaker under certain obligations, directive insofar it is an invitation to trust and declarative insofar it marks a formal reconciliation and new beginning. Why then should apologies be understood *solely* through the lens of expressive speech acts? And is there reason to extend that doubt to other so-called expressive speech acts as well? However, both an extended investigation into whether apologies are expressive speech acts and a critical reevaluation of the fivefold schema are outside the scope of this thesis.
[27] Their case bears similarities to Searle's (1976) regarding intentional states, where intentional states such as being sorry, confusion, doubt, fear and so on can be expressed in terms of more basic intentional states such as beliefs, desires and intentions.

than saying that one regrets what one did. One can regret something in private when no one else is concerned, such as regretting getting caught. But apologies express a particular kind of regret, the kind of regret that comes with a change of heart. An analysis of regret is not an exhaustive analysis of being sorry. Guiraud et al. use their formalized account of emotions underlying expressive speech acts to define the psychological state of *being sorry* as an agent expressing to a hearer regret that the hearer did not reach a goal (2012: 1035-1037). But this is the sorry we use in condolences, not apologies. It contains that which the notion of regret seemed to lack: an acknowledgement of the injustice that has incurred the other. But this definition of sorry fails to place the speaker's own agency at the heart of this injustice. The notions of regret and being sorry are unequivocally inseparable in apologies, as referring to a propositional content in an apology implies that one relates to it regretfully and the apology is deemed insincere if the speaker is not really sorry for what they have done. A true juxtaposition of these two terms deserve a chapter on their own, but here it will suffice to say that a particular kind of regret is an element of being sorry but not vice versa.

But we are none the wiser about the psychological state implied in an apology. Words such as 'sorrow', 'being sorry', 'regret', 'remorse' and 'repentance' are all typically associated with apologies, and the only definitions so far are Guiraud's et al. definitions of regret and being sorry. But what do these concepts *mean* and what are their relation? Again, we are stuck with asking for definitions. Speakers of ordinary language have little trouble in fluently using these concepts, but the *know-that* involved in articulating what they mean does not come with the *know-how* of using them. Both to say one is sorry and being able to point out when others' usage of this word is off are tacit skills we possess as ordinary language speakers. But when is it that we have the right to say that someone's use of the words 'I am sorry' is somehow off? For instance, a friend who says they are sorry for what they did, only to follow it up with a "but…" and defend their course of action, is a friend who we call out for just *saying* that they are sorry when they are in fact not. There are, so to speak, correct and incorrect uses of words which we are tacitly familiar with. But whether we are able to articulate it or not, there must be some conditions which must be met for the usage to count as correct or incorrect. To create such an analysis of all the terms involved is a lengthy matter, far outside the scope of this thesis. However, for the sake of focusing on apologies, analysing the meaning of 'being sorry' will cover much of the distance. After all, the notion of 'sorry' when expressed in apologies can be made to cover for the notions of regret and sorrow as well, although it cannot do so for instances of regret and sorrow outside the scope of apologies. The task at hand is to articulate when and where resorting to the word such as 'sorry' is called for, when its meaning becomes stretched or abused, or when there is no doubt that the word has been correctly understood.

So why not use the definitions given by Guiraud et al.? While reducing 'being sorry' to a matter of beliefs and intentions raises some concerns, that is not main point of contention here. The main

chafing point lies between the bottom-up approach of this thesis and their linguistic top-down approach. They are two different approaches to the *meaning* of a concept, in this case the concepts of apology and what it means to be sorry. The efforts by Guiraud et al. to understand individual expressives by locating their place in an abstract schema was set in motion by Searle (1969) when he concluded that *an* expressive speech has the point of expressing *a* psychological state. It follows quite naturally from this assumption that expressive speech acts ought to be understood by making those underlying psychological states neatly align with corresponding expressive speech acts, expressed in terms of beliefs and desires. I argue that because their starting point is that of expressives *as a whole,* it forces particular speech acts boxes that do not correspond to the speech acts as we know them in ordinary language. Such an approach fails to explain the meaning of being sorry in a way that resembles the way we teach each other as part of our form of life – sometimes successfully and sometimes not – what it means to be sorry. If an analysis of apologies is to bear any resemblance to the way apologies "do work" in everyday speech, then that analysis must be as familiar with its failures – mistakes, misunderstandings, hyperextensions, insincerities, abuses of language, to mention a few – as its successes. Otherwise, how do we know what is at stake when really meaning one is sorry and really meaning what one says? To learn what it means to really be sorry is a long process that is as much an ethical journey as it is one of being a competent speaker. It is not enough to "feel" the right thing: one must learn to view things in a certain light at will if one wish to be a competent apologizer. Reducing sorry to a matter of beliefs and desires – or worse, to basal emotions – downplays the long and arduous road that some of us still have ahead of us in becoming competent in this sense. Being true to the way we understand 'being sorry' entails that our definition, when pressed, does not have to retreat back to the abstract schema as its foundation. It must retreat to the examples which we as ordinary language speakers can come to see as reflecting our forms of life.

2.5.1 Being sorry as cognitive attitude

Many of the misleading simplifications regarding 'being sorry' can be reduced to unclarities regarding psychological states, especially when it comes to expressive speech acts. In Austin's work, this category is referred to as 'thoughts or feelings' (1962); in Searle's work (1969), they are psychological states; Norrick (1972) and Guiraud et al. (2011) refer to them as emotions; McMyler (2012), when discussing Austin's *Other Minds*, refer to them as mental states. I will argue next that if one seeks to stay true to what Austin meant by "thoughts and feelings", being sorry is best described as a *cognitive attitude*. Cognitive attitude has traditionally been used to denote beliefs, doubts and other epistemic relations to propositional content, but in this chapter, I seek to expand the concept in order apply it to speech acts more generally. The main reasons for favouring cognitive attitudes over psychological states are twofold: first, it best captures the way in which being sorry relates the speaker *to* something and tells us *how* they

relate to it; second, it does not discriminate between thoughts, feelings, opinions, stances, attitudes, vibes, takes, relationships and other nouns which denote a kind of stance that a person can have to a propositional content. Cognitive attitude highlights what Searle referred to as the intentionality of the mind, i.e. our mind's capacity to be *about* something (Searle 1983). The relation a person holds to a propositional content is a matter of the mind *doing* something insofar it means assuming a certain stance. It serves to remind us that although we see things in one light, we could see them in another.[28] To view these the holding of these cognitive attitudes as a matter of doing something also properly appreciates the way in which our thoughts and feelings cannot ever be properly separated: as many before me have pointed out, emotions are forms of thought (e.g. Solomon 1993, Damasio 1994). Let us explore these arguments in order, which will also elaborate the way in which being sorry is a case of holding (rather than having or being in) a certain cognitive attitude.

Cognitive attitude expresses the way a person thinks and feels as a matter of relating to something: speech acts in general, and expressives in particular, express the way the speaker relates to the propositional content of that speech act. For instance, a claim implies a kind of cognitive attitude, namely that of a belief. To congratulate someone implies that one perceives something as being in the favour of whoever one is congratulating. To lament is to relate to whatever one laments over with a certain sorrow or melancholy, perhaps framing the object of lament as being a nuisance or grievance. An accusation is to settle on some link between the accused and what they are accused of. The examples are numerous, and all speech acts at least seem at first glance to express some kind of stance that the speaker takes in relation to whatever it is they are speaking about. When speaking of "psychological states" instead of "cognitive attitudes", the propositional content referred to by that psychological state is hard to separate from the psychological state itself. After all, psychological state does not in itself indicate any separation between the propositional content and the particular way the mind relates to it. Being sorry about one thing is one psychological state, while being sorry about another thing is another psychological state. However, when speaking of cognitive attitudes, it is easier to properly distinguish between what it is the speaker relates to and how it is that they relate to it. It is thus in line with Searle's notion of intentionality, which is the way our minds are *about* something. Remorse, which he pinpointed as the cognitive attitude expressed in apologies (interchangeably with regret, it seems), is one of such intentional states, in the sense that remorse must be *about* something (Searle 1983: 4). If someone says they are feeling remorseful and one asks them what it is they are regretting, the answer "oh I just feel remorseful in general" is misplaced in a sense that "feeling a bit down in general" is not. One may not have a clear view of what

---

[28] Our perceptions are judgments, subject to many of the same criticisms and expectations as our actions do: to express some unsavoury opinion, for example, can subjected to moral reproach, because expression of an opinion is itself a kind of action. To view something in a certain light cannot (although we often try) be defended as being "simply" a perception that follows "naturally" from whatever it is that is being perceived in that light.

an alternative outcome may have been, but one implies that one views one's current situation in a certain remorseful light, regardless of how it is they frame that situation. Same can be said about intentions and beliefs, as one does not intend or believe in general, but rather relate to various propositional contents as a matter of intending or believing.

Speaking of people's being sorry as a matter of "psychological state" assumes there to be a mind which lies in a particular state. Sincerity is then a matter of the person having their mind in the particular state that they express or imply via speech acts. If the state of mind expressed and the actual state of mind are in correspondence, then the speech act is sincere. But first of all, what is this mind, independent from the mental states that it might find itself? And second, what is this mental state that also already contains what that mental state is *about*? Cognitive attitudes, on the other hand, avoids committing to the former superfluous distinction while making the latter necessary distinction. This allows for the proper appreciation of how attitudes towards propositional contents can change while still counting as the same cognitive attitude across various examples, as well as the propositional content staying the same despite our various ways of relating to it. When speaking of psychological state, neither the object nor relation cannot be changed without entailing a qualitative change in the psychological state as a whole. This undermines the continuity of the mind across both attitudes and content. In fact, our language relies on our ability to maintain such continuities across change: for example, one's use of the word 'regret' for both having a cup of coffee too much and having broken a promise are different, yet also share what it means to regret something. It is only against the backdrop of their similarity that we can make fine distinctions between these two kinds of regret. Because the term 'psychological state' does not properly separate the *way* of relating and the *propositional content* related to, it also fails to account for this continuity.

The term 'psychological state' also comes with significant cartesian baggage that seeks to distinguish thought from emotion, without even addressing other ways of relating to that which our speak and think *about*. As mentioned above, cognitive attitude captures the various ways in which we relate to things without discriminating between thoughts, feelings, attitudes, opinions, takes, vibes, stances and so on. Even the words 'think' and 'feel' are ambiguous: what Descartes had in mind when saying that "he thinks" is vastly different from what we mean when we ask our loved ones "what do you think?" upon trying on a new piece of clothing. We are asking for how they see it, judge it, feel about it, draw associations from it, vibe with it, whether they can meet it halfway, work with it or they refuse to do so. What they think *can* of course be expressed as a matter of cartesian rational thought, but it is not the only kind of response (perhaps not even the appropriate one) that will count as an answer. Their answer will involve part feeling, part thought, part taste, part bias. What we want from them is to express that relation in a way that goes beyond just being a feeling, thought, opinion, bias and so on. We confide in their judgement and adjust our own accordingly.

As another example one may speak of doubt as being a cognitive attitude. One thinks of that which one doubts in a certain way, but also can feel that doubt, it can amount to seeing something in a certain light, it can be a performative stance taken towards that which one doubts, or giving the benefit of doubt as a matter of decency, and so on. But doubt does not have to be one of these at any given time, as it can be an amalgam of them all, with some of these descriptions perhaps hitting home while others miss the mark. Most often we do not *simply* think or *simply* feel when we speak of something: we do a bit of both when we see things in a certain light. The dichotomy between thought and emotion is itself burdensome outside of special circumstances where we consider that difference to be imperative. Thoughts, feelings and the rest are all ways in which we describe our value laden judgements just like a wine taster describes how they relate to the wine they are tasting. It is at the junction of 'cognitive attitudes' that we find thoughts, feelings, perceptions, opinions, biases, stances, attitudes, vibes and so on as being wrapped into one. As we shall see, reducing the cognitive attitude of 'being sorry' to something as simple as a particular feeling or thought is misleading. When we want to make sure that someone is truly sorry for what they did, we ask of them to *see* what they did in a certain light. If they do not do so already, we plead of them to adjust their point of view and see the matter in that light, to at least entertain the thought or simply show some compassion for whomever it is that deems themselves worthy of an apology. It is precisely because our point of view can change that we can come to have a change of heart; a change which allows us to once again see the matter through the eyes of the other (or, ultimately, refuse to do so).

Lastly, the notion of cognitive attitudes highlights the way we *hold* those attitudes, rather than simply *have* them. When speaking of psychological states, the person having those psychological states is too easily rendered as being passively awashed by the fancies of their mind. The difference between having and holding is that in the latter we can find the component of agency and all the notions of responsibility and accountability that follows, while in the former that sense of agency is amiss. Having a vindictive feeling bars criticism and praise in a way that holding a grudge does not. To speak of cognitive attitude, however, highlights the way that how we decide to view things really is a matter of doing: to see something in a certain way can change, if we make the effort to see things in another light. One must speak of holding a grudge as a cognitive attitude rather than a psychological state if one wants to properly highlight the way we can be expected to drop it as a matter of growing up or be commended for standing up for ourselves. Being sorry, I would argue, is much like that of holding a grudge in the sense that it is not a matter of feeling, but a matter of trying to view the past in a certain light. Without the ability to speak about frames of mind as something we bear at least some responsibility for, we cannot speak of minds as being appropriate for a situation. Without that sense of responsibility, we cannot articulate what it is that we require of each other when we hope, or demand, of that we are sorry for what we have done.

2.6 What apologizing commits us to

An apology is a matter of expressing something more than an emotion or valence of mind. It is also something more than an expression of a speaker's beliefs and desires. By apologizing, the speaker undertakes the commitment that they indeed see matters in a certain light. If the speaker's commitment is recognized by the hearer, they enter a bond of trust.

Next, the commitments that are undertaken in a proper apology are articulated along with the ways one can come to go back on those commitments. The best way of doing so is by exploring the kinds of utterances which, when following the utterance "I am sorry", result in a contradiction similar to the one featured in Moore's Paradox. In Moore's Paradox, to assert something entails a belief in the propositional content of her claim and the paradox emerges when one explicitly denies the belief without denying the assertion. The speaker's belief in the propositional content of the assertion is implicit to asserting just like being sorry is implied in apologizing. But how does one *know* if the person *actually* believes what they are saying? Here, we stumble on the issue of sincerity, but in the context of assertions instead of apologies. In order to know if the person is committed to having holding belief they implied in their assertion, one must look at the commitments that a speaker undertakes in making an assertion. One must look for the contradictions between their words, by which they undertake commitments, and actions, which either line up with or contradict those commitments. For instance, if a speaker says their favourite football team will win tomorrow, then that speaker is implying that they believe their team will win. Their belief, which is a cognitive attitude they hold about tomorrow's game and their team's chance of winning, can be described as a kind of commitment. If they were to then make a bet, they would be bet on that team; if that team were to lose, they would be in disbelief. If the speaker were to instead bet on the other team or rationalize that their team chances of winning were low, then they would be contradicting what was implied in their prediction, namely that they had a certain belief about their team's chances of winning. We would doubt *their word* that they ever really believed in their victory. In making an assertion, they undertook certain commitments that are tied to the belief that was implied in that speech act.

Notice that doubting the speaker's word about their beliefs does not make references to anything mental that they have privileged access to (and we do not). One does not make the claim that one can peer into someone's mind like a telepath is imagined doing. Rather, one is simply listening to what they say and using one's eyes to measure those words against their actions. Whether the retort "you don't get to say what I believe and don't believe" dismisses what belief is assumed to follow from their claim or what actions contradict the belief that was assumed to follow, it is not so much wrong as it is beside the point; it expresses a misunderstanding of how beliefs are understood in ordinary circumstances. Their

claim implied their commitment to a belief and that committing to that belief entailed certain restrictions to their actions. To go outside of those restrictions is to contradict oneself, or in other words, to go back on one's words. It is by their future actions that they will come to bolster or undermine the *lived* truth of their words. Again, what restrictions follow from which words is a matter of convention, something which cannot be pinned down here because conventions change from culture to culture and time to time. In order to show that one understands what words entail, one must speak and act in certain convention bound ways, such as acting as if one believes what one says when one asserts something to be the case (Searle 1969: 48).

The sincerity condition of assertions is that one holds a certain belief and the sincerity condition of apologies is that one is sorry for what one apologises for. The one who is sincere in their apology is the one who implies that they see things in a certain way and acts accordingly. If being sorry describes a certain cognitive attitude towards that which one says one to be sorry about, what kind of stance is being taken by the speaker? Since we do not have "direct access" to the contents of other minds (whatever that would entail), we must rely on the internal relation between saying and doing. We must judge others on when they have used their words correctly, when the meaning of those words have been stretched, when they are abused for personal gain and so on. The task at hand is to describe what it means to be sorry: what commitments are undertaken in "saying so" and how ought one to "act accordingly"? If successful, one should be able to outline the kinds of actions that are *contradictory* with the point of view committed to in the apology.

Note that what is implied in an apology is something we are already familiar with as speakers of ordinary language (the kind of familiarity highlighted by Cavell (1976) when he speaks about becoming a native language speaker). It is this familiarity which allows us not only to apologize, but also live out our lives in ways that are coherent with how we profess to view things. In fact, these two are one and the same: to apologize *is* to undertake certain commitments and living by them. We already know, tacitly, what we are committing ourselves to when making use of our privilege of being understood, trusted, relied upon as an informant, being able to repudiate and challenge, accuse and atone, declare and stipulate, and so on. It is because we know what consequences our words have when uttered against the background of a common language that we are able to partake in a common world of both language and ethics. Furthermore, knowing the consequences of our words is what allows us to recognize when someone has fallen back on their words and when to accuse of others of having said they are sorry when they are in fact not.

But articulating this tacit knowledge is a challenge. And that challenge is further aggravated by the fact that our mastery of language does not mean we do not make mistakes. It is precisely in those

shortcomings that out mastery is put to use as we show others how they have 'gotten hold of the wrong end of the stick' when it comes to a word or when that word has been put to the wrong use.

Since being sorry is a cognitive attitude and expressing cognitive attitudes imply certain commitments, contradictions occur when one apologizes yet acts in ways that undermine others' trust in the fact that one view things the way implied in the apology. At times, this may happen because of a misunderstanding. A person might not have a proper grasp of what it *means* to be sorry and does not quite understand the relation that word bears to viewing things in a certain way. Because they do not understand the precise relation between the word and a particular point of view, they do not know what kinds of actions are contradictory to saying that they are sorry. Or, the meaning of the word may be hyperextended: one might extend the notion of responsibility, agency and causality in a way that results in apologizing for silly things. These kinds of mistakes are the ones we see in children who are being introduced into the language game of apologizing: they may say that they are sorry without having the faintest idea what this means beyond being able to avoid further reproach. Calling it an abuse of language is to go too far however, as abuse implies that one knows the proper use of a word. In actual abuses of language, words are being purposefully misused, stretched or obfuscated. One may feign ignorance regarding the implications one makes in speaking or one may misconstrue one's commitments or break them on purpose. In order to properly abuse language, such as saying one is sorry in order to fool someone back into an unhealthy relationship, one must already have a grasp of how the other will understand it when relying on *common convention*. A child is subjected to different expectations when using the word 'sorry' because they are still learning what that common convention is.

The time has come to articulate the apology's equivalents to Moore's Paradox. Let us use the following example. *S* apologizes to *H* for action *p* by saying "I am sorry that *p*". Note that even if S said "I apologize for *p*", the apology implies that the speaker is sorry that *p*, just as an assertion implies a belief. Being sorry is a certain kind of cognitive attitude, but in order to characterize what kind of relation it is that *S* has to *p* in saying that they are sorry that *p*, we must ask what kinds of commitments *S* is undertaking in saying that they are sorry that *p* and what subsequent actions constitute a contradiction of those commitments.

At least four commitments are being undertaken by *S* in saying that she is sorry that *p*.

2.6.1 Injustice

1.) The apology implies that *S* recognizes that *p* constitutes some form of injustice towards *H*. In other words, *S* implies a certain kind of judgmental stance towards the propositional content of the apology. What meets the definition of injustice of course varies with culture, situation and people. Likewise, the general theory of justice guiding the definition of injustice will vary, just as the understanding of

responsibility, agency and causality can vary from culture to culture. But albeit the content of the word justice and injustice is culturally variant, every culture has *some* conception of justice. Agreement on matters of justice is one of the building blocks of having a culture. Thus, one should here only ascribe a minimal definition as to what is being implied by judging something as an injustice. Justice and injustice correspond at their most basic level to which states of affairs we are committed to support, enact or realize and which states of affairs we are committed to judge, hinder or undo. To say *p* is an injustice implies at the very least that one does not perceive *p* as something that is desirable or something *H* deserved.

Let us entertain an example of this contradiction at work. Imagine *S* telling *H* that they are sorry for snitching on him to their employer for unionizing. That apology may very well be taken by *H* as having been sincere and a bond of trust becomes reestablished between the two. One of the reasons that *H* comes to trust *S* is that *S* has implied in her apology that she perceived her actions as counting as an injustice towards him; it was something she should not have done. It might very well be that both silence and snitching entailed some kind of injustice, in one case being an injustice towards their workplace and in the other being an injustice towards a fellow comrade. But in apologizing to *H* for having snitched on him, she at least implied that she now considers silence to have been the lesser of two evils. If she now were to pledge during a workplace meeting that she is committed to combating any unionizing on the workplace as a matter of being *the right thing to do*, *H* would be right to ask, "was your apology insincere or is your pledge insincere?" After all, she is now implying in her pledge that snitching on him *was* the right thing to do and that he *did* deserve it: this entails a contradiction between saying that she views her actions in one light and then acting as if she viewed them in another.

Another story will exemplify this point. Imagine two friends, *S* and *H*, the latter suffering from alcoholism and the other being aware of his affliction. *H* the alcoholic has been banned from the local pub and asks *S* if she can bring him along as an avec the next time she heads out to a house party. She agrees. But as he is brought along, he does not end up at a party. Rather, she has organized a public intervention where friends and family has been brought together to wish him luck on his recovery. Furious at this betrayal, *H* makes it clear that he expects *S* to apologize for having betrayed him like this. She apologizes, saying that "I am sorry, *but* it is for your best". Unsurprisingly, *H* rejects her apology. But the reason for doing so is not because she isn't right about the fact that it is the best for him. He is in need of help and should not be brought along to a house party. It is not even the fact that he has been publicly outed as an alcoholic that irks him. The real reason he is hurt is because she betrayed his trust and lied to him. In saying that she is sorry, but she did what she thought was best for him, she is on the one hand inviting him to trust her that she really is sorry. On the other hand, she implies that doing what is best for *H* outweighs the betrayal of his trust. There is no contradiction in saying that betraying *H*'s

trust is an injustice while considering that what one did was still justifiable. It is a contradiction, however, for her to say that she is sorry for betraying his trust while at the same time not acting as if she regretted it. The utterance becomes an excuse for her actions, rather than an apology.

It is because of this that the utterances such as "I am sorry, *but* you deserved it" and "I am sorry, *but* it was for your own good" are non-apologies. It entails a contradiction in the Moorean sense because it explicitly denies something that was implied in the apology, namely that S considers p to count as an injustice towards H. Likewise, to say that one is sorry, only to then act *as if* one did not perceive the matter as an injustice, entails a contradiction.

2.6.2 Agency

2.) The apology implies that S believes there was an alternative course of action. In other words, p was not inevitable. These alternative courses of action are referred to as r. This is not to say that there is a *specific* action r that was the right alternative. After all, we often regret our actions without having a clear idea what we should have done instead. But recognizing there to have been alternatives in any sense of the word means that the speaker does not consider their action inevitable. To say that there was an alternative to p implies at the very least that the speaker admits there to have been a choice, without subscribing to having a full picture of their alternatives. This sense of agency is imperative to apologies, because it allows the speaker to assume responsibility for their actions.

Imagine the following case. S and H are colleagues who attend an important meeting. S apologizes for having picked a rather late bus which had the consequence of making her arrive late to the meeting. In doing so, she implies that she did have some other course of action available to her, such as taking an earlier bus. If not, then she would instead defend her course of action by saying that she took the earliest bus available and that there was nothing more to be done. If her colleague insisted on S apologizing, her other colleagues might back her up: "you don't owe her an apology for scheduling a meeting at a time where you were unavailable!" But perhaps she did have an alternative, such as calling in early to reschedule the meeting. If the meeting needed to be rescheduled and she did try to call in order to move it, there would again not be anything for her to apologize for: "You don't owe her an apology for *her* not answering the phone when you did try to reschedule!" But perhaps there was an even earlier alternative, such as when she *promised* she would be there on time. If she promised she would be there on time, but the earliest bus was too late and the meeting could not be rescheduled, then her alternative course of action would have been to *not* promise being there in the first place. Thus, S adjusts her apology: she says she is sorry for promising things she could not keep, but she does not apologize for the bus she chose or for not having tried to reschedule the meeting. The latter things are beyond her control, but the

act of promising was her choice. She locates the point of responsibility that is appropriate to what she is apologizing for.

There are indeed situations where we feel inclined to apologize for that which we do not control. Apologizing for the actions of others is one case among many, barring the cases where one's responsibility to stop them from hurting someone outweighs their responsibility for not doing it (such as when overseeing one's children or caring for someone with a mental disability). But apologies can be out of place when one attempts to apologize for that which is inevitable or, at least, is the best possible outcome. Imagine the unlucky person doomed to pull the lever in Foot's trolley problem. He pulls the lever and one person is ran over by the trolley instead of five. He did pull the lever, so the blood of that person is on his hands. There is no denying that it *was* his choice to pull the lever. The victim's family reproaches him for having robbed them of a beloved family member. What would he say in his apology? As the trolley hurtled down the tracks, his alternatives were to not pull the lever, resulting in the death of five people, or to pull the lever, killing one. Pulling the lever was the best possible option and while he certainly did pull the lever, a better option was not given to him. Instead of defending his actions as being (according to him) the rational thing to do, imagine that he attempted to apologize for his actions. If he were to say "I am sorry for having killed your son, but it was the best option available", would be truly be apologizing? There is a dimension on inevitability to his situation, and while he did have at least one option, it is arguably worse than what he chose. Was it really an option at all? In fact, he is part of those who should be consoled for having been involved, against his will, in such a twisted scenario. But apologizing yet defending his actions as being the only option he had is a more subtle form of the same Moorean paradox as above.

Another example is needed to fully illustrate this point. Imagine that S is known for being very outspoken about what she thinks. Despite H being aware of this personality quirk (or flaw of character, as he might see it), he ends up being very hurt by something that S says to him in her fervour of honesty. The purpose was, in this case, not to purposefully hurt his feelings, but to simply be honest. She apologizes to him, but says that, in her defence, that's just who she is. She might say "sorry for saying that, *but* I just couldn't help speaking my mind" or "sorry for saying that, *but* you know I am like that so you shouldn't take offense". Something has gone awry in her apology; a sentence that started out as an apology quickly turned into that of making excuses. A contradiction has occurred because her apology implies that she takes responsibility for p (or at least subscribes to a common view of what it is we can be held responsible for) while at the same time explicitly denying that responsibility. Rather than admit that choosing her words carefully is her responsibility, she shifts the blame unto H for taking offence. This is not to say that she has to apologize: it might be admirable to be honest and straightforward. But then one must ask the purpose of saying that she is sorry while at the same time defending her actions as

inevitable. What is the purpose of apologizing when she does not perceive herself to have anything to apologize for? Is the purpose to get *H* off their back? To shift blame somewhere else? To immunize themselves against criticism or deprive others of the chance to criticize her for not having apologized? Or, perhaps, the purpose is merely to avoid conflict? Whatever the case may be, if her apology includes a denial of responsibility for the propositional content of the apology, it results in a non-apology.

It is because of this that utterances such as "I am sorry, *but* I did not have a choice" and "I am sorry, *but* I just couldn't help it" are non-apologies. It entails a contradiction in the Moorean sense because it explicitly denies something that was implied in the apology, namely that *S* believes there to have been an alternative to the course of action she is sorry for. Likewise, to say that one is sorry, only to then act *as if* one did not have control over the actions one apologized for, entails a contradiction.

2.6.3 Normativity

3.) In apologizing, S admits seeing the difference between *p* and *r*. In other words, *p* and *r* are admitted by *S* as not being interchangeable. What is to be considered a meaningful difference varies, of course, from case to case. To illustrate the kind of difference at stake, consider how one may acquire something by ordering and threatening a person: the fact that both are a case of receiving what one wants by telling the other to give it is not to say that they are morally interchangeable. In saying that she is sorry, *S* admits that the alternative course of action available to her actually was different in some meaningful way.

Imagine the following case. *H* asks *S* of a favour: *S* is to prepare a surprise birthday party for their mutual friend, *T*. When *H* arrives with *T* to the party, they notice that *S* has done the bare minimum. *S* apologizes for the fact that she has not organized the party that *H* had envisioned. But rather than end the apology there, she claims that she really doesn't see the difference anyway since "it's still a birthday party all the same" and "it's not like T is going to care". Because of what *S* has said, *H* draws the conclusion that *S* is in fact not sorry for having botched the birthday party. Rather, she seems to deem his reproach pedantic, putting it outside of the realm of things that warrant an apology. He therefore judges her apology to be insincere: if she was truly sorry, then she would *recognize* that meeting the minimum criteria of *H*'s request and actually putting effort into *T*'s birthday party are, in fact, not satisfied by one and the same action. *H* reserves the right to reproach *S* for being insincere, since in saying that she is sorry for not having thrown *T* a nice enough party she implies that she recognizes the difference between what she has done and what she ought to have done.

Let us consider another example. *H* has asked *S* to repay the debt she owes him. The due date for paying the debt passes, and *S* has yet to repay her debt. Upon calling *S*, she makes it known *H* that she did not realize he was in such a hurry to get back his money. When *S* finally repays her debt, she apologizes for having taken so long to do so. If she were to insist in her apology that she still does not

see why he cares so much on receiving the money immediately, then some form of contradiction would occur: she would be apologizing yet not admitting that paying her debt on time is in some sense different from respecting the due date. She bites her tongue and simply apologizes for being late with the payment. *H* accepts her apology and says that all is well; "just make sure to repay on time the next time". After all, the due date had been set on the assumption that *H* would have the debt repaid in full by the due date. The reason *H* accepts the apology by *S* is that he trusts her to know the difference between repaying in time and being late on her payment. After all, her apology implied as much. At a later point in time, *S* borrows money from *H* again. Just as last time, the due date passes, and her payment is late. Upon calling *S*, she once again expresses confusion over the fact that he is in such a hurry to have the debt repaid. As a defence for her payment being late, she says that a day or two should not make much of a difference. But this time a contradiction occurs, since she had already implied in her earlier apology to recognize the difference between a debt paid in time and a payment that is overdue.

It is because of this that utterances such as "I am sorry, *but* I don't really see what difference it makes" and "I am sorry, *but* it was just once" are non-apologies. It entails a contradiction in the Moorean sense because it explicitly denies something that was implied in the apology, namely that there is a significant difference between what they did and what they ought to have done. Likewise, to say one is sorry, only to then act *as if* one did not see any meaningful difference between what one did and what one ought to have done, entails a contradiction.

2.6.4 Intention

4.) The apology implies the *intent* of avoiding *p* in the future in favour of *r*. This implication is perhaps already a part of 1-3 above insofar declaring something an injustice itself implies a commitment to avoiding it in the future, as well as admitting what one did as not being that which one ought to have done implies that one prefers to do the preferable thing in the future. Nevertheless, it is worth reiterating as a separate point that being sorry about something implies the intention to avoid it in the future, especially when apologies contain promises, such as when one apologizes for an action *by* promising not to do it again.

Imagine the following case. *S* and *H* attend a conference together after co-authoring a paper together. Despite suffering from debilitating stage fright, *H* is able to present the paper if *S* is also there to present it with him. However, during the conference *S* is presented with the opportunity to also present the paper at the same time in a neighbouring university. Seeing this as an opportunity to bring their research some much needed visibility, *S* accepts the offer and leaves *H* to present the paper on his own. No matter how the presenting of the paper went on either part, *H* reproaches *S* for having abandoned him like that. As *S* apologizes for having abandoned him despite having agreed to present the paper

together, *H* tells her to promise that she will never put her work before her colleagues again. She replies that she cannot promise such a thing, because she does not intend to put her colleagues before her work in the future either. In other words, *S* says that she is sorry for abandoning *H* yet denies the intention to avoid doing so in the future. A contradiction occurs between the commitment that is implied to have been undertaken in the apology (that she intends to avoid the apprehensible action in the future) and her explicit denial of intending to change her ways.

Let us consider another example. Imagine a person *S* apologizing to *H* for having cheated on him in a relationship. She convinces *H* that she truly regrets her actions and that she promises to never do it again. *H* accepts her apology, and although their relationship will never be the same, they undertake a rocky journey towards reestablishing a sense of trust. Imagine, however, that *S* is in a chat group with her friends. *H* wonders what she talks about with her friends in said group, but *S* assures that it is nothing he needs to know about. It is, after all confidential between her and her friends. Although the temptation is strong, he resists the urge to read their chat log unbeknownst to her. One day, however, one of *S*'s friends sends *H* a screenshot of the chat log. In that screenshot, *H* can see that *S* has told her friends that *H* is a sucker for believing her apology and that she is would still like to see the person she cheated on *H* with. *H* is, understandably, hurt and furious. But his feelings are not hurt because she was ready to cheat on him *again*, as if it would have been a second instance of the same injustice. Rather, he is hurt because she betrayed his trust that she was committed to their relationship. Thus, their bond is once again soured as her earlier expression of regret has been rendered null and void.

One more example is needed to illustrate this point. *S* apologizes to *H* for having secretly been reading in his diary, despite *H* explicitly forbidding it. The matter is put behind them until *S* is caught reading his diary again. This time, *H* does not simply reproach *S* for having read in his diary but having done so *again*. Note that his second reproach is not just a repetition of the first, but rather contains a different propositional content altogether: *H* reproaches *S* for having apologized yet done it again, which is not the same as having simply done something reproachable twice. Now, her previous apology has soured, having been retroactively rendered void by her repeated infractions. Perhaps she truly meant what she said in the moment. Upon apologizing, her intention might indeed have been not to read his diary again. But it is in conflicts like these where we do not necessarily doubt a person's benevolence. Rather, we doubt whether she has understood what it means to say she is sorry for having done so. People make mistakes – they are often doomed to repeat them as well – but after a certain amount of reiterations of infraction, reproach and apology, we start to doubt whether the person in question understands what she commits herself to by saying that she is sorry for what she has done.

It is because of this that utterances such as "I am sorry, *but* I will not change my ways" and "I am sorry, but I don't intend to stop" are non-apologies. It entails a contradiction in the Moorean sense

because it explicitly denies something that was implied in the apology, namely that *S* intends to avoid *p* in the future. Likewise, to say one is sorry, only to then act *as if* one still was prepared to do *p,* entails a contradiction.

3 Conclusion

What words mean is a matter of convention – rules for what counts as what and what follows from it – and this applies to when we speak of how we think or feel too. To say that one is sorry counts as implying a recognition of responsibility the same way as saying it is raining implies one holds that belief. It is also only against the background of conventions that we can understand what it means to apologize correctly or incorrectly, or to invoke the notions of sincerity and insincerity when judging others. And it is because there are different components to these conventions that apologies can fail in various interesting ways.

Many of those conventions are about the commitments which we undertake as we speak. For instance, to say one promises comes to count as undertaking a certain commitment whether one intended for it to do so or not. However, I argue that the notion of undertaking commitments is integral to all speech: we acquire the privilege to make speak meaningfully by understanding how what we say limits what we can do if we wish for what was said to not be undermined by what we subsequently did. If I say I hold a certain belief, then some of my subsequent actions are limited; to go outside those limitations is, in the eyes of others, to go back on my word. Likewise, if I say I am sorry, then too are my subsequent actions in some sense limited if I wish for my apology to hold.

To undertake those commitments is to extend a hand to the other as an invitation to trust. Although trust is quite obviously an integral part to making a promise, I argue that the notion of trust is integral to our everyday speech about how we think and feel. The notion of trust is invoked here to avoid falling into accounts of human intersubjectivity (or the lack thereof) which I not only find deeply distressing, but in which we cannot quite see ourselves saying that we are sorry.

The sincerity condition, which is integral to the success of an apology, posed us with the problem of we can come know whether a person is being sincere. Without some direct access to minds, contradictions between word and mind cannot be known. Instead of construing sincerity as a correspondence between word and mind, I sought to show how speaking makes the speaker undertake commitments. This allowed the construal of sincerity as a correspondence between word and action, which allows the understanding of apologies without straying too far from Austin's philosophy. By applying Moore's paradox to apologies, four possible contradictions between word and action were discovered, which entails four commitments a speaker undertakes in saying that they are sorry.

First, apologizing implies viewing the propositional content of one's apology as an injustice to the person one apologizes to. This injustice, in turn, implies that one does not seek to defend one's

actions as being justified. This commitment can be contradicted by actions which imply that the speaker sees their action as being justified. The hearer trusted the speaker that they saw their action as not justified.

Second, apologizing implies a recognition of agency that places the propositional content within one's sphere of responsibility. This responsibility, in turn, implies some form of agency: the outcome was brought about by the speaker's actions and the speaker acknowledges that they could have acted otherwise. This commitment can be contradicted by actions which imply that the speaker could not help bringing about the outcome. The hearer trusted the speaker that they carried responsibility.

Third, apologizing implies that one sees the meaningful difference between what one did and what one perhaps ought to have done. This differentiation implies that the speaker can locate the point of contention in her actions. This commitment can be contradicted by actions which imply that the speaker does not see alternatives courses of action as making a meaningful difference. The hearer trusted the speaker that they could understand what the problem with their previous behavior was.

Fourth, apologizing implies that the speaker does not intend to repeat their problematic behavior. This intention implies that the speaker chooses to refrain from carrying out that same action in the future. This commitment can be contradicted by actions which imply that the speaker does in fact no hold this intention to refrain from repeating the problematic behavior in the future. The hearer trusted the speaker that they would avoid repeating the behavior that was deemed problematic. These four, although the product of our investigation, have always been known to us as ordinary language speakers. What was accomplished in this thesis was first the delineation of conceptual tools with which it became possible to articulate what it means to say that one is sorry in a way that resembles the way in which those words are learned in their ordinary language setting.

What about the introductory example, then? If I were to travel back in time and be my own interlocutor, how ought (or could) I have responded to my past self when invited to prove a claim about insincerity? Perhaps I would have taken my past self for their word when they said they were sorry, despite my better judgement: "Fine, then. I cannot know if you mean what you say. It remains to be seen." Given that sincerity must be understood as a correspondence between word and action instead of word and mind, such a reply would not be entirely out of place. But there is something to be judged in the caustic actions of my past self, some abuse of language that ought to be diagnosed. Perhaps the apology, being a take-it-or-leave-it offer, was best to be declined, lest one was ready to be hurt again: "I do not trust you because you are being insincere and you appear insincere because I do not trust you." The apology ordinarily functions as a gesture of reconciliation, something which forms bonds of trust between people, but it was now being deployed by my past self as a means of cutting one's losses via the invocation of skepticism, at the cost of severing those bonds. With no invitation to trust and no mutual recognition of commitments being undertaken, there could be no saying one is sorry and meaning it.

Oskar Lindblom
FHPT vid Åbo Akademi

Bibliography

Austin, J.L., 1962, How To Do Things With Words, Oxford University Press, London

Baz, A., 2011, Knowing Knowing (that Such and Such), *The Philosophy of J.L. Austin* (ed. Martin Gustafsson & Richard Sorli), pp. 146–174, Oxford University Press, New York

Baz, A. 2012, When words are called for : a defence of ordinary language philosophy, Harvard University Press, Cambridge, Massachusetts

Blum-Kulka, S. & Olshtain E., 1984, Requests and apologies : a cross-cultural study of speech act realization patterns (CCSARP), *Applied Linguistics*, Vol. 5, Issue 3, Autumn 1984, pp. 196–213, Oxford University Press, Oxford

Cavell, S. (1976), Must We Mean What We Say? Cambridge University Press, Cambridge

Damasio, A., 1994, Descartes' Error : Emotion, Reason, and the Human Brain, Putnam Publishing, New York

Falkenberg, G., 1990, Searle on sincerity, *Speech acts, meaning and intentionality* (ed. Armin Burkhardt), pp. 129–146, Walter de Gruyter, New York

Fotion, N., 2001, John Searle, Princeton University Press, New Jersey

Gustafsson, M., 2005, Perfect Pitch and Austinian Examples: Cavell, McDowell, Wittgenstein, and the Philosophical Significance of Ordinary Language, *Inquiry* Vol. 48, No. 4, pp. 356-389, August 2005, Routledge, Oslo

Hertzberg, L., 2013, Norman Malcolm ja käypä kielenkäyttö, *Tavallisen kielen filosofia* (ed. Joose Järvenkylä & Ilmari Kortelainen), pp. 133–147, Gaudeamus OY, Helsinki

Holmes, J., 1990, Apologies in New Zealand English, *Language in Society*, vol. 19, No. 2 pp. 155-199, Cambridge University Press, Cambridge

Järvenkylä, J., 2013, Austinin puheaktiteoria ja kielellinen merkitys, *Tavallisen kielen filosofia* (ed. Joose Järvenkylä & Ilmari Kortelainen), pp. 77–89, Gaudeamus OY, Helsinki

Kovalainen, H., 2013, Stanley Cavellin tavallisen kielenfilosofian ylitys, *Tavallisen kielen filosofia* (ed. Joose Järvenkylä & Ilmari Kortelainen), pp. 148–162, Gaudeamus OY, Helsinki

Lagerspetz, O., 1996, The tacit demand : a study in trust, Filosofiska Institutionen, Åbo Akademis tryckeri, Åbo

Lasinkaité, L., 2013, The importance of being earnest : the true art of apology (M.A. thesis), Lithuanian University of Educational Sciences, Faculty of Foreign Languages, Vilnius

McDowell, J., 1998, Mind, Value and Reality, Harvard University Press, Cambridge, Massachusetts

McMyler, B., 2011, Believing what the Man Says about his own Feelings, *The Philosophy of J.L. Austin* (ed. Martin Gustafsson & Richard Sorli)., pp. 114–145, Oxford University Press, New York

Murphy, J., 2014, Apologies in the discourse of politicians : a pragmatic approach, University of Manchester Faculty of Humanities, Manchester

Norrick, N. R., 1978, Expressive Illocutionary Acts, *Journal of Pragmatics*, Vol. 2, pp. 277–291, Elsevier, Amsterdam

Ryle, G., 2009, The Concept of Mind (60th Anniversary Edition), Routledge, Abingdon, New York

Schedler, G., 2007, Should There Be an Apology for American Slavery?, *Public Affairs Quarterly*, Vol. 21, No. 2, April 2007, pp. 125–148, University of Illinois Press, Champaign, Illinois

Searle, J. 1969, Speech acts : an essay in the philosophy of language, Cambridge University Press, New York

Searle, J. 1979, Expression and meaning : studies in the theory of speech acts, Cambridge University Press, New York

Searle, J. 1983, Intentionality : an essay in the philosophy of mind, Cambridge University Press, New York

Searle, J. & Vanderveken, D., 1985, Foundations of Illocutionary Logic, Cambridge University Press, New York

Solomon, R. C., 1993, The Passions: Emotions and the Meaning of Life, Hackett Publishing Company, Inc., Indianapolis, Indiana

Wittgenstein, L. 1992, Filosofiska Undersökningar (transl. Anders Wedberg), Bokförlaget Thales, Stockholm

5. Svensk sammanfattning

Ursäkter (dvs att *be om ursäkt* för något och inte att *ge ursäkter* för något) en talhandlingar där begreppet *uppriktighet* framstår som särskilt viktigt. Trots att löften är löften oberoende om man menar det eller inte, är det inte lika klart om ursäkter räknas ifall man betraktas som att inte mena det man säger. Detta beror på att ett av villkoren för lyckad talhandling är det s.k. uppriktighetsvillkoret. När talhandlingen *att be om ursäkt* analyseras, framstår detta uppriktighetsvillkor som bärande på filosofiska problem. Ett av orsakerna till dessa problem ligger i den definition av uppriktighet som används i talhandlingsfilosofisk och lingvistisk litteratur: uppriktighet är en korrespondens mellan ord och sinne. Talhandlingen är uppriktig ifall talarens sinnestillstånd motsvarar det hon påstår i samband med talhandlingen. Denna definition lyfter fram klassiska filosofiska problem gällande medvetande och berättigande för påståenden som görs om medvetande: hur vet åhöraren ifall talhandlingen är uppriktig? På hurudan grund avvisar hon ursäkter när hon finner dem ouppriktiga? Och på hurudan grund kan hon bli övertygad om att talaren är uppriktig?

Men sådana påståenden är en väsentlig del av s.k. vardagsspråk där talare verkar kunna göra påståenden om sinnestillstånd och åhörare verkar smidigt avgöra ifall dessa påståenden är sanna eller falska. Avvisandet av en ursäkt som ouppriktig är ett exempel på detta. Frågan är då hur människor i en vardagsspråkskontext gör sådana påståenden och omdömen. Eftersom majoriteten av forskning i ursäkter har utförts inom lingvistisk pragmatik och inte filosofi har dessa filosofiska problem inte behandlats i litteraturen i det specifika sammanhanget av att be om ursäkt. Trots att filosofisk litteratur har till viss utsträckning behandlat uppriktighetsvillkoret (t.ex. Falkenberg 2012), har den inte behandlats på ett sätt som kan lätt överföras till att beröra uppriktighetsvillkorets roll i ursäkter. Denna avhandling är inte menad för att besvara dessa problem, utan att visa hur denna problematiska definition av uppriktighet och hur de kan kringgås genom ett perspektivskift.

För att kringgå de filosofiska problem som uppstår p.g.a. definitionen av uppriktighet, skiftas definitionen från en korrespondens mellan ord och sinne till en korrespondens mellan ord och handlingar. Detta placerar uppriktighetsbegreppet i sitt vardagsspråkliga och etiska sammanhang, vilket är ett sammanhang där begrepp som tillit, plikter, åtaganden och integritet hör hemma. Uppriktighetsbegrepp kan alltså förankras i hur talare betraktas ta sig an åtaganden som förpliktar dem till att handla på vissa sätt, samt vilka handlingar som är oeniga med dessa åtaganden. M.a.o. frågas det på vilka sätt talare binder sig då de till exempel ber om ursäkt. Den relation som skapas mellan talare då den ena talar uppriktigt grundar sig alltså inte på kunskap om att personens sinne är i överensstämmelse med hennes ord, utan på tillit om att hon kommer att handla i överensstämmelse med dem. Denna korrespondens mellan ord och handling gäller även då talaren hävdar sig inta ett visst förhållaningssätt:

huruvida hon menar det hon säger då hon påstår sig förhålla till hennes tidigare handlingar är en fråga om integritet mellan hennes ord och handlingar, inte hennes ord och sinne.

När man ber om ursäkt utförs i synnerhet fyra åtaganden. För det första åtar man en viss förhållningssätt gällande handlingens rättvisa. För det andra medger talaren en viss relation av agens mellan talare och handling. För det tredje förhåller sig talaren till handlingen normativt. För det fjärde antyder talaren en avsikt gällande framtida handlingar, nämligen att undvika utföra samma handling i framtiden som hon ber om ursäkt för. Alla fyra åtaganden medför begränsningar på talarens handlingar och överskridande av dessa begränsningar underminerar talarens integritet, dvs meningsfullheten av hennes ord.