

Social Spheres: logic, ranking, and subordination

Gillian Russell

Draft from October 2022

forthcoming in a volume of *Minnesota Studies in the Philosophy of Science* on feminist logic,
edited by Audrey Yap and Roy Cook. Please cite official version once available.

This paper uses logic—a formal language with models and a consequence relation—to think about the social and political topics of subordination and subordinative speech. I take *subordination* to be a matter of three things: i) ranking one person or a group of people below others, ii) depriving the lower-ranked of rights, and iii) permitting others to discriminate against them.¹ Subordinative speech is speech—utterances in contexts—which subordinates.

Section 1 introduces the topic of subordination using examples from the 1979 novel *Kindred* by Octavia Butler (Butler, 2004). Section 2 uses these examples to clarify and illustrate the definitions of subordination and subordinative speech. Sections 3 and 4 then develop a way of modeling subordination using a *system of social spheres*, an adaptation of (Lewis, 1973)’s approach to modeling the relation of comparative similarity on worlds for counterfactuals. Section 4 looks at three possible applications for this work: giving truth-conditions for social quantifiers, identifying fallacies involving such expressions, and explaining the pragmatics of subordinative speech. The last section anticipates objections and raises further questions.

1 Pragmatics for Time-Travelers

Dana Franklin, a young black writer from 1970s Los Angeles, travels back in time—spontaneously and against her will—to a Maryland plantation in the early 1800s: a slave state in the time of slavery. There she meets white slave owners: the Weylins and

¹I take this explication of subordination from the literature and particularly Langton (1993) and Maitra (2012). The restriction to people is unnecessary for my purposes, but they are my focus here. (Russell, 2019, §7.21) One question that I cannot answer to my satisfaction is what the relationship between the three clauses should be. Must each be satisfied for subordination to occur, or is satisfaction of just one sufficient? For the purposes of this paper I’ll assume that each needs to be satisfied, but I stipulate this more as a matter of convenience than conviction. I am attracted to two alternative possibilities. On one, subordination is merely a matter of rank, but reductions in rank tend—in fact—to be accompanied by changes in what one is permitted to do and how others are permitted to treat one, i.e. i) is the grounds for ii) and iii) and so is often accompanied by them. On the other possibility, social status *consists* in social power, where this is cashed out in terms of social permissions, so that what it is to be lowered in rank is to have ones permissions restricted, and others’ permissions with respect to one increased.

their son Rufus.² Over the course of several trips to the past she meets the Weylins' black slaves, as well as several free black people. On one occasion Dana's white husband, Kevin, accompanies her on her hazardous travels.

As is usual with time-travelers, Dana and Kevin find themselves in conversations with people whose assumptions are quite different from their own. Initially, Dana does not know what has happened and assumes that she is still in California, whereas the young Rufus Weylin assumes she is in Maryland. Others assume that Dana is a man (she is wearing trousers) or that she can't read.

Sometimes Dana, with growing awareness, alters her backstory in order to avoid conflict with other people's assumptions; she claims to be from New York instead of California, since she knows that California was still a Spanish colony, and explains away her clothing by saying that Kevin (who they assume to be her owner) has not bought her anything gender-appropriate—an explanation the slaves find convincing.

At other times Dana tries to correct the assumptions of her interlocutors. Early in the novel, Rufus tells us that his mother thought Dana was "Just a strange n—."³ Dana disputes the assumption that this is an acceptable way to refer to black people:

"What's wrong?" he asked. "Why are you mad?"
"Your mother always call black people n—, Rufe?"
"Sure, except when she has company. Why not?"

His air of innocent questioning confused me. Either he really didn't know what he was saying, or he had a career waiting in Hollywood. Whichever it was, he wasn't going to go on saying it to me.

"I'm a black woman, Rufe. If you have to call me something other than my name, that's it."
"But..."

I want to focus on three kinds of assumption on which Dana differs from her interlocutors. These are assumptions about:

- i) Dana's social status
- ii) what Dana is permitted to do
- ii) what others are permitted to do *to* her

²Rufus turns out to be one of Dana's own ancestors, and her time-traveling to happen whenever Rufus is in mortal danger. She comes to suspect that the purpose of her trips is to ensure her own survival by keeping Rufus alive.

³I won't use or quote the n-word in this paper, though it appears in Butler's novel. Academic norms conflict on the appropriateness of quoting slurs, with some arguing that it is inappropriate to treat a slur as special, as I do here by replacing all but the first letter. For my part, I give significant weight to the argument of a black student who said she encounters this word in her ordinary life, and doesn't need to see it in an academic paper to be reminded of its full force, as well as to the arguments of a recent unpublished paper by DeeAnn Spicer. I accept that if you are the target of this slur, it can be especially exhausting and distracting to read it over and over again while working or studying.

Some of the examples of subordination in this paper retain a distracting brutality anyway. I have tried to minimise gratuitous use of examples while also avoiding the opposing risks of appearing to encourage neutrality, or implying that subordinative harm is generally minor or trivial.

i) is a bit more complicated than the others, so I'll start with ii). Dana and Kevin find that people in 1815 have different assumptions about what Dana is permitted to do.

“Who're you?” asked Rufus.
“My name's Kevin—Kevin Franklin.”
“Does Dana belong to you now?”
“In a way,” said Kevin. “She's my wife.”
“Wife?” Rufus squealed.
I sighed. “Kevin, I think we'd better demote me. In this time...”
“N—s can't marry white people!” said Rufus. (61)

Other characters assume that Dana is not permitted to go where she likes, say what she likes, read when she wants to, or share a meal with a white person. They expect her not to do what Mr Weylin and Rufus refer to as “talking back.”⁴ More generally, Mr Weylin assumes that the only conversational contributions Dana is permitted to make are the ones he demands:

“Yes, I was reading to him,” I admitted. “He asked me to do that too. I think he was bored lying in there with nothing to do.”
“I didn't ask what you thought,” he said. (94)

Moving on to iii)—things people are permitted to do *to* Dana—Rufus assumes, as we saw, that people can refer to her using the N-word, and that his father can have her whipped. Rufus' mother assumes she can give Dana orders, insult her, and strike her with impunity. All the temporally local characters assume that Kevin may sell Dana.

Coming finally to assumptions of type i)—those about relative social rank—I find that this category is more difficult to illustrate, even though it seems obvious that Dana's social rank is lower than that of many white people, because they are rarely made explicit in conversation.⁵ I conjecture that explicit claims about relative status are avoided because such assumptions generally make both those of high and those of low status uncomfortable.⁶ Still, people's assumptions about Dana's status are naturally understood as the *source* of assumptions of kinds ii) and iii). It is because of people's assumptions about Dana's status relative to Mr Weylin that they assume that she isn't permitted to talk back to him, and that he *is* permitted to have her punished. Still, there are many conversations in which social hierarchy simmers below the surface, where people are implicitly ranked on the basis of colour, on the basis of being a slave or free, and on the basis of their assigned tasks (e.g. working in the house vs working in the fields.)

⁴Talking back to his father is also something that Rufus is whipped for towards the beginning of the novel “What exactly are you supposed to be sorry for?” I asked when Weylin was gone. I spoke very softly. “Talking back,” said Rufus. He thinks everything I say is talking back. So I don't say very much to him.” (109)

⁵One arguable exception to this is on page 78 when Sarah (one of the household slaves) explains that she was permitted to keep her fourth child because “... bless God, Carrie ain't worth much as the others cause she can't talk. People think she ain't got good sense.”

⁶This is certainly sometimes true, perhaps because an explicit representation of someone as having a certain status can make salient the question of whether that status is deserved.

(Jake to Dana) “You think you’re white!” he muttered. “You don’t know your place any better than a wild animal.” (180)

(Sarah to Dana) “You ain’t no field n—, but you still a n—. Marse Rufe can get mad and make things mighty hard for you.” (164)

(Mr Weylin to Rufus): “You ought to be ashamed of yourself! A n— can read better than you!” (108)

Assumptions about status are especially close to the surface when Dana and Rufus argue about what to call each other. Dana, as we’ve seen, initially protests being referred to with the n-word. Rufus finds this puzzling.

Dana: “I don’t like that word, remember? Trying calling me black or Negro or even colored.”

Rufus: “What’s the use of saying all that?” (61)

But when Rufus and Dana first meet, he also has some ideas about how *he* should be addressed by her:

“And you don’t call me ‘Master’ either.”

I surprised myself by laughing. “Master?”

“You’re supposed to.” He was very serious. “You want me to call you black.” (25)

Rufus thinks wanting to be called ‘black’ is an *equivalent* request to wanting to be called ‘Master.’ This is a bit like thinking that if you agree to call someone ‘human’, they have to agree to call you ‘Boss’ or ‘Sensei’ or ‘Gov’nor’. The requests aren’t equivalent, *unless* both are merely demands to be referred to appropriately, something that makes sense only on the assumption that one person has higher status than the other.

To sum up: among the assumptions that the characters in *Kindred* make in conversation with Dana are assumptions about what she is allowed to do, and assumptions about what other people are allowed to do to her. A special case of these involves speech that is or isn’t permitted: talking back (not permitted to her) and calling her the the N-word (permitted to others.) Finally, the characters also make assumptions about Dana’s status, and the best evidence we have for this is that it explains their judgements about what it is permitted and what is not.

2 Subordination

These three kinds of assumption correspond to the three features of subordination identified in (Langton, 1993, 302–3). On Langton’s approach, to subordinate someone is to:

- i) rank them below others
- ii) deprive them of rights

iii) legitimate discrimination against them.

In this section I am going to say a little about each of these clauses, and I'll use incidents from *Kindred* to illustrate.⁷

The kind of *ranking* involved in the first clause is best understood as a *social* ranking. In calling the ranking *social*, I mean to imply that it is neither a ranking by objective, intrinsic worth, nor a personal, subjective ranking determined by the opinion of a single person. To see why, consider Mr Weylin and his slave, Luke. In the sense of *ranking* relevant for subordination, Mr Weylin is ranked above Luke. But this ranking is independent of any relative objective worths they might have; whether Mr Weylin is (counterfactually) the natural, God-given master of the lowly Luke, or whether Luke is the nobler, richer soul, donating his labour to the venous old Weylin,⁸ there is a ranking relationship which holds in virtue of various social facts. Luke is, let us say, listed as Mr Weylin's property on the county tax records, as having been purchased by him on such-and-such a date, and they both live in a society that embodies and enforces certain practices and rules: slaves are not permitted to retire, or move to Canada, and their owners are entitled to give them orders, collect the fruits of their labour for free, sell their children for profit, and have them punished. This social situation makes a great deal of difference to Luke's life quite independently of any moral or objective ranking of worths that holds between them. But it also holds independently of any one person's subjective ranking of the two individuals, determined only by preferences or psychology (much as we might rank our favourite restaurants or movies). You and your friends can be mistaken about your subordination (as Kevin initially is with Dana), but this isn't sufficient to undo it. Changing hearts and minds would not—on its own—be enough to undo subordination, you also need changes in social practices, including norms, laws, and enforcement.

Moving on to ii) and iii), it can be helpful to think of both these clauses as concerning *permissions*. Depriving people of rights is about making things impermissible that would be permissible were the targets not subordinated, whereas legitimating discrimination concerns granting permissions *to others* to discriminate. Again, think of these permissions as *social* permissions, rather than moral, objective, or purely subjective. For example, Luke is clearly *morally* permitted to run away. Still, his society would punish him for this. He does not have social permission to run away, and he cannot run away *with impunity*—without risking social punishment. Mr Weylin is socially permitted to punish or sell Luke, and so he can do either with impunity. Since Mr Weylin cannot sell or punish just anyone in his society, this is a way in which he has social permission to discriminate against Luke. Just as social permissions hold regardless of the genuine moral permissions, so they also hold independently of most people's personal subjective estimation of the permissions. In a painful development in *Kindred*, Luke gives Dana his personal—somewhat liberating—take on how a slave may behave: "Don't tell them 'no.' Don't let them see you mad. Just say 'yes, sir.'"

⁷Though I take my start from Langton, I don't know whether she would endorse everything I go on to say about these three clauses here. My project is not merely Langton-exposition, but rather: I think I recognise the phenomenon of subordination she identifies, and I have more to say about it that will be helpful for what I do later in this paper.

⁸Or even if ownership relations between persons are metaphysically impossible, so that, objectively speaking, no-one can be in a master-slave relationship.

Then go 'head and do what you want to. Might have to take a whippin' for it later on, but if you want it bad enough, the whippin' won't matter much." Chillingly, it turns out that Luke's personal estimation of what a slave could do differed from the social reality: when Dana next returns, Mr Weylin has sold Luke to a trader who plans to ship him to New Orleans (likely a death sentence.) Rufus explains that Luke "didn't show much sense." and that she had better "show some sense yourself, Dana." "Luke...he would just go ahead and do what he wanted no matter what Daddy said. Daddy always said he thought he was white. One day, maybe two years after you left, Daddy got tired of it. New Orleans trader came through and Daddy said it would be better to sell Luke than to whip him until he ran away."

Some social permissions may be encoded and enforced as laws, but they need not be so encoded: for example, Dana is not permitted to read, rest, or talk when she likes.

Assuming the understanding of subordination above, we might call an *action* "subordinating" if it does one of several related things. It might make someone subordinate who was not subordinate before (e.g. overthrowing a monarch and making them subordinate to the new one.) Or instead it might merely increase a degree of a subordination that was already present (as in Germany's decision to revoke citizenship from Jews in the 1930s—in this case Jews were already subordinated, and the action subordinated them more.) Or finally, the action might just be one more element reinforcing subordinative social norms that are antecedently in place, even if it does not *increase* the degree of subordination, (e.g. a 1930s US school principal who always demands that women teachers resign when they become engaged to be married. The principal's actions might not subordinate women any more than they are already subordinated by the standing social norm of the marriage bar, yet his actions nonetheless subordinate since they are a part of the mechanism that maintains women's subordinate position.)

Dana is subordinated as a result of her time-travel to 1800s Maryland; her social rank is reduced.⁹ But many others who are local to that time are the subjects of subordinating actions—even when the actions don't increase the degree of their subordination—because they still reinforce the subordination: e.g. Mr Weylin's forcing slaves to watch other slaves being whipped, selling slave children, and beating Rufus for "talking back."

Subordinative speech is speech which subordinates. As with other actions, speech subordinates if it makes someone subordinate who wasn't before, increases the degree of their subordination, or reinforces their antecedently subordinate position. Plausibly, the announcement from the UK Medical Register that Andrew Wakefield had been struck off was a subordinating speech act: it lowered his rank, revoked his permission to practice medicine in the UK, and permitted others to discriminate against him, e.g. when trying to hire doctors. Langton's famous example of the Apartheid-era South African Legislator who enacts new legislation with the words "blacks are no longer permitted to vote" is a case in which the degree of subordination is increased: the utterance reduces its targets social status, deprives them of the right to vote, and legitimates discriminatory behaviour by election officials. Finally speech can be an action which reinforces existing subordination: examples might include telling sexist jokes in an

⁹If 1970s LA had been a socially egalitarian society, then we would have been able to say she was subordinated in the first sense. But of course it wasn't.

antecedently sexist society, or Rufus' saying in a tone of outrage, "N—s can't marry white people!"

3 Ranking Worlds

In David Lewis' book *Counterfactuals* (1973), the actual world is just one among many worlds, getting along in modal space. Call this space W , and its inhabitants $w_0, w_1, w_2...$ etc. Each world w is mapped by a function $\$$ to a set of *spheres of accessibility*, $\$w$. Lewis calls $\$$ a *centered system of spheres* if and only if it meets these four conditions:

- C) Centering: $\{w\} \in \$w$
- 1) Nesting: for $T, S \in \$w$, either $S \subseteq T$ or $T \subseteq S$
- 2) Closure under unions: if $\bigcup S$ is the union of a set of spheres in $\$w$, then $\bigcup S \in \$w$.
- 3) Closure under (non-empty) intersections: if $\bigcap S$ is the non-empty intersection of a set of spheres in $\$w$, then $\bigcap S \in \$w$.

The conditions are justified by considering the fact that the set of spheres is meant to carry information about the *comparative overall similarity* of worlds: the smaller a sphere around a world w , the more similar to w a world has to be to be a member. We might say that the system of spheres *ranks* the worlds by similarity to w .

Here is how we justify each constraint: Centering is justified because each world is maximally similar to itself, and more similar to itself than any other world. Nesting is justified because if we were to have a violation, we would have spheres $T, S \in \$w$ such that neither $S \subseteq T$ nor $T \subseteq S$, i.e. there is $w_1 \in S$ but $\notin T$ and $w_2 \in T$ but $\notin S$. In that case, we'd have both that w_1 is more similar to w , but also that w_2 is more similar to w (this would be a problem given that the spheres are meant to carry information about *overall comparative similarity*); the nesting condition prevents this. For unions, suppose that $w_1 \in \bigcup S$ while $w_2 \notin \bigcup S$. Then w_1 but not w_2 must be a member of some sphere in S . Hence w_1 is more similar to w than w_2 , and hence $\bigcup S$ should be a sphere around w .¹⁰ Intersections is justified in the analogous way: suppose that $w_1 \in \bigcap S$ but $w_2 \notin \bigcap S$. Then there is some sphere in S such that w_1 is in it but w_2 is not, and hence w_1 is more similar to w than w_2 is. Hence $\bigcap S$ should itself be a sphere.

4 Social Spheres

Each of us is just one among many persons, getting along in social space. Call this space U (for *universe*), and its inhabitants, $p_0, p_1, p_2, ...$ etc. Suppose that the people are organised into a set of social spheres, R , conveying information about *social rank*. The smaller the spheres of which an individual is a member, the higher their rank.

One difference between similarity and social rank is that social rank need not be relative to a third person; we can think of *being ranked lower than* as a binary relation

¹⁰Closure under unions and intersections are superfluous where there are only finitely many worlds around w .

holding between a person p_1 and a person p_2 , rather than as a ternary relation holding between a world w_1 and a world w_2 relative to a “target of similarity” w .^{11,12} The result is that we need not think of our system of spheres as a function from a person, but can simply regard it as a fixed set of social spheres, R , defined on the domain of persons, U .

The set of spheres represents social rank: if p_1 is a member of a sphere that p_2 is not a member of, then $p_2 < p_1$ (p_2 is ranked below p_1 .) On this way of modeling things each sphere is, as it were, an exclusive social club, and if your friend is a member but you are not, then your friend ranks more highly than you do.

What constraints on spheres are justified, given the goal of conveying information about overall social rank? First, note that we do not require a centering condition and, in the absence of a ranking-target (an Empress) one would not be well-defined. The rank of p_2 below p_1 will thus be absolute, not relative to a highest-ranked person. But the remaining Ludovician constraints can be adopted and justified in a way that is analogous to the similarity case: Nesting: if sphere $S, T \in R$, then either $S \supseteq T$ or $T \supseteq S$. For suppose not: then there is some $p_j \in T$ who is not in S , and some $p_k \in S$ who is not in T . But this would represent the unhappy situation in which p_j is ranked overall higher than p_k (because it is a member of a sphere that p_k is not a member of) and also vice versa. Unions: suppose that $p_1 \in \bigcup S$ while $p_2 \notin \bigcup S$. Then for some sphere in S , p_1 but not p_2 must be a member. Hence p_1 is ranked above p_2 , and hence $\bigcup S$ should be a sphere in R . Finally, non-empty intersections: suppose that $p_1 \in \bigcap S$ while $p_2 \notin \bigcap S$. Then p_1 is a member of some sphere in S that p_2 is not a member of, and hence ranks above p_2 . Hence $\bigcap S$ should be a sphere in R too.

Pulling together what we have so far: let R be a set of subsets of the domain of persons, U . We call R a *set of social spheres* if meets the following three conditions:

- 1) Nested: for $T, S \in R$, either $S \subseteq T$ or $T \subseteq S$
- 2) Unions: if $\bigcup S$ is the union of a set of spheres in the system R , then $\bigcup S \in R$.
- 3) Intersections: if $\bigcap S$ is a non-empty intersection of a set of spheres in the system R , then $\bigcap S \in R$.

Within Lewis’ framework, we might call something a *modal frame* if it is a pair $(W, \$)$ in which W is a non-empty set of worlds and $\$$ a centered system of spheres. Analogously, call something a *social frame* if it is a pair (U, R) , in which U is a non-empty set of individuals, and R a set of social spheres. A modal frame becomes modal *model* with the addition of an interpretation function, which assigns appropriate values to the primitive non-logical expressions relative to each $w \in W$: $(W, \$, I)$. A subsequent definition of truth-in-a-modal-model for sentences of a language L will allow us to

¹¹I don’t want to insist that social rank can *never* be treated as a ternary relation on people. Perhaps there is a society where we can think of rank as social distance from the Empress. Then we could think that p_1 is higher ranked than person p_2 iff p_1 is *socially closer to the Empress than* p_2 . This would be treating social rank as a ternary relation. But since we won’t in general want to assume that every society has a highest ranking person, we’ll simplify things here by treating the ranking relation as a binary relation.

¹²The ternary relation might be more useful for representing a *narcisist* society in which every person considers themselves the highest ranking person and ranks all others by social distance from themselves. In practice I think that’s not how social ranking actually works. Not everyone thinks that they are at the top.

define a relation of (modal) logical consequence on L : A is a logical consequence of Γ , ($\Gamma \vDash A$) just in case every model in which (every member of) Γ is true is a model in which A is true.

Analogously, a social frame becomes social *model* with the addition of an interpretation function, which assigns appropriate values to the primitive non-logical expressions: (U, R, I) . A subsequent definition of truth-in-a-social-model for sentences of a language L will give us a relation of (modal) logical consequence on L :

Definition ((Social) Logical Consequence). *A sentence A is a (social) logical consequence of a set of sentences Γ , ($\Gamma \vDash A$) just in case every social model in which (every member of) Γ is true is a social model in which A is true.*

That's a logic, or at least a recipe for one (as we will see in the next section there are some important details to iron out.) Is it a feminist logic? It is rather more general than that, since the ranked social spheres needn't represent genders; they could be races, socio-economic groups, age groups, military ranks etc. But I think it's fair to say that it is a social logic, or logic of social subordination—in the same sense that a logic defined over the centered systems of spheres can be a modal logic—and moreover that it is a logic that could be of special interest to feminist philosophers.

5 Applications

We arrived at our sets of social spheres by adapting Lewis' way of thinking about similarity. But Lewis' picture is not *merely* a way to think about similarity, it serves various roles and makes different things possible. It can be used to:

1. give truth-conditions for modal claims like $\Box A$ and $\Diamond A$;
2. give truth-conditions for counterfactual claims like $A \Box \rightarrow B$ and $A \Diamond \rightarrow B$;
3. identify counterfactual fallacies, such as strengthening the antecedent, transitivity and contraposition;
4. explain the pragmatics of counterfactual assertions.

Could social ranking models have related applications? In the remainder of this paper I take Lewis' four applications in turn and ask whether anything similar can be done with socially ranked models.

5.1 Social Quantifiers

Start simple. Lewis' systems of spheres can provide truth-conditions for ordinary unary modal operators like \Box and \Diamond . We might say:

$\Box A$ is true at a world w in a model $M (= \langle W, \$, I \rangle)$ just in case A is true throughout w 's sphere of accessibility.

But *which* sphere of accessibility? w is at the center of a nested *set* of spheres in M . We can entertain various approaches here. At one point in *Counterfactuals* Lewis suggests

that different necessity operators are associated with different spheres (5) and that suggests that there might be a set of such operators, perhaps one for each sphere. These might have different subscripts e.g. \Box_{ph} for when we're quantifying over physically possible worlds, \Box_{no} for when we're quantifying over nomologically possible ones, etc.

Alternatively, we might have a single necessity operator whose associated sphere is determined by conversational context.¹³

$\Box A$ is true at a world w (in a model $M = \langle W, \$, I \rangle$) in a context c just in case A is true throughout the sphere of accessibility determined by the model and c .

Absent a context, it will be unclear which sphere of accessibility is to be used. But contexts—which may contain information gleaned from surrounding sentences—can contribute information which narrows the options. For example, a speaker might add more information:

- (1) a) It's impossible for you to get to New York by 5pm; you've got that dinner to go to in Carrboro.
- b) It's impossible for you to get to New York by 5pm; all the flights are cancelled.
- c) It's impossible for you to get to New York by 5pm; the fastest any extant vehicle can go is 2200 mph.
- d) It's impossible for you to get to New York by 5pm; it's 5.01pm now.

A natural way to interpret the 'impossible' in 1a) is as restricted to spheres of worlds in which you meet your current social obligations. For the one in 1b) it is as restricted to spheres of worlds in which you take an ordinary commercial flight to New York, for 1c) it is as restricted to worlds in which we have our current technologies, and for 1d) to those in which practical time travel is impossible for human beings. If we were to insist on interpreting the 'impossible' in 1a) as quantifying over a much larger sphere—say one which included all metaphysically possible worlds where you have a dinner to go to in Carrboro—then the sentence would be false, because there are worlds where you go to New York and get there for 5pm even though you had a dinner to go to in Carrboro (i.e. ones where you miss that dinner.) Here some of Lewis' other work on *accommodation* is relevant: in cooperative conversational contexts, where a claim would only be true if we were presupposing p in that context, an utterance of that claim tends to lead (all things being equal) to that presupposition being added to the context. So if I utter 1a), and interpreting me as saying something true requires restricting to a narrower sphere, then we restrict to a narrower sphere. All things being equal, we try to interpret the speaker in a way that makes what they say true. (Lewis, 1979) One job for a modal model then, is providing a sequence of domains for the interpretation of context-sensitive operators, like \Box .

¹³Or at least, an operator which can be associated with more than one sphere of accessibility. I mean to remain uncommitted on whether there will be multiple necessity-style operators, e.g. one for alethic, one for deontic, or something of that sort.

A social model can do something similar, though the details must be different. One key difference is that while systems of spheres rank worlds, social spheres rank individuals. Worlds are things relative to which sentences get truth values, but individuals are not; *snow is white* is true in some worlds, false in others, but it's not true *for some people* and false for others.¹⁴ The model-theoretic analogue for the truth of a sentence when the object is an individual, rather than a world, is the satisfaction of a 1-place predicate (an open formula with one free-variable.) Predicates *are* satisfied by some individuals and not by others. And the syntactically appropriate expression for quantifying over individuals isn't a sentential operator, like \Box , but a unary first-order *quantifier*: a device that, syntactically, attaches to 1-place predicates to form sentences, and, model theoretically, to sets to give truth-values.

Thus the truth-clause for a social universal quantifier might be:

$\forall \alpha A$ is true in a model, $M = \langle U, R, I \rangle$, and context c , just in case A is satisfied by **every** member of the social sphere $S \in R$ determined by c .¹⁵

Other monadic quantifiers can also be defined, including an existential quantifier and 'most':

$\exists \alpha A$ is true in a model, $M = \langle U, R, I \rangle$, just in case A is satisfied by **at least** one member of the social sphere $S \in R$ determined by M and c .

Most αA is true in a model, $M = \langle U, R, I \rangle$, just in case $\text{card}|A|_S > \text{card}|U - A|_S$, where S is the sphere determined by M and c .¹⁶

What's distinctive of the quantifiers we are looking at, given their sensitivity to ranking, is a certain asymmetry in favour of higher ranked individuals: higher ranked individuals must be taken into consideration even if we are talking about lower ranked individuals, but if we are mostly focusing on higher ranked people, then we might not be including lower ranked people in the domain of the quantifier. They are analogous to "further out" possibilities. Such quantifiers could be used to interpret natural language sentences like:

(2) Most people have seen *Hamilton* now.

This is something that someone might say—especially someone who goes to see a lot of Broadway musicals. *Hamilton* was a successful musical and has now toured many cities. Still, the truth is that the tickets cost hundreds of dollars and price and location

¹⁴Perhaps some special sentences—like *pickled jalapeños are delicious*—get person-relative truth-values, but that is not the general case.

¹⁵As a reminder of our stipulations from section 4: U is a non-empty domain of individuals, R a set of social spheres obeying the constraints from section 4, and I an interpretation function for the primitive non-logical expressions. Basically this is a first-order "Tarski" model plus a set of spheres. In the most interesting cases A and B will be an open formulas with α free in each.

¹⁶Here $|A|$ is the set of individuals in U satisfying the formula A and $|A|_S$ is the intersection of that set with the sphere S (the restriction of $|A|$ to S). $\text{card}(X)$ denotes the cardinality of a set X . So $\text{card}|A|_S > \text{card}|U - A|_S$ means, roughly, that there are more "As" in S than "non-As", i.e. that *most* individuals in S "are As".

remain significant barriers to most people getting to see it, so that in fact (quantifying broadly) most people have not seen the musical. In order to interpret the speaker in such a way that their utterance comes out true we have to assume that the domain of quantification is a narrower social sphere—that is, that the domain of quantification is restricted to individuals of higher rank. Other examples of sentences which could be interpreted using such quantifiers include:

- (3) (a) I’m not going to get stuck in this job; I’m going to be **someone**. (Even though people who get stuck in the job are also ‘someone’ if we quantify more broadly.)
- (b) You really have to be **someone** to be asked to Buckingham Palace. (They aren’t counting the people who are asked to interview for jobs as cleaners.)
- (c) **Nobody** needs to see another *Nutcracker/Carmen/Spiderman*. (Here the quantification is restricted at least to people who are old enough to be well-versed in many classics.)
- (d) **Everyone** who lives at the Manor goes up to London for the season. (Even though many of the servants live at the Manor and remain there.)
- (e) **Nobody** in this city can manage without a cleaner/personal trainer/nanny/driver/life-coach. (Though many of the cleaners/personal trainers/nannies/drivers and life-coaches do manage without.)
- (f) **Everyone** summered in the Hamptons. (Used to describe 1950s NYC, in which the wealthy left the city in the hottest summer months.)
- (g) **Everyone** should own a gun. (Used by racist NRA supporters who don’t mean to include members of the Black Panthers.)
- (h) **All** men are created equal. (Written by a slave-owning Jefferson on behalf of a country in which slavery was legal.)
- i) “Many men of course became extremely rich, but this was perfectly natural and nothing to be ashamed of because **no one** was really poor—at least no one worth speaking of.” (Adams (1978))

5.2 Variable Binary Social Quantifiers

The primary application of Lewis’ systems of spheres is not to give truth-conditions for the monadic modal operators \Box and \Diamond however, but to give truth-conditions for counterfactuals. Counterfactuals are *binary* modal operators; syntactically, they take two sentential arguments to form a sentence and, model-theoretically, their truth-conditions are given in terms of the truth-values of those two component sentences, as follows:

$A \Box \rightarrow B$ is true at a world w (in a model $M = (W, \$, I)$) if and only if either

1. no A -world belongs to any sphere S in $\$,w$,¹⁷ or
2. some sphere, S , in $\$,w$ contains at least one A -world, and $A \rightarrow B$ holds at every world in S .

As unary operators are to binary operators, so unary quantifiers are to binary quantifiers.¹⁸ Binary quantifiers take two formulas (usually each has a single variable free) as arguments, and their satisfaction conditions are given in terms of the satisfaction conditions of those component formulas. We can illustrate with the binary universal quantifier: \forall^2 . The syntactic rules for \forall^2 say that if A and B are formulas, and α is an individual variable, then $\forall^2\alpha(A, B)$ is a formula. Following a tradition in linguistics, we will call the formula in the A -position the *restrictor*, and the formula in the B -position the *scope*. The set of individuals satisfying the restrictor formula, $|A|$, is then the *restrictor set* and the set of individuals satisfying the scope formula, $|B|$, is the *scope set*. $\forall^2\alpha(A, B)$ is true iff $|A| \subseteq |B|$, i.e. just in case the scope set is a subset of the restrictor set.

For example, $\forall^2x(Fx, Gx)$ will be true just in case all the things which “are F” (members of $|Fx|$) also “are G” (members of $|Gx|$.) Other binary quantifiers will require different relations between the scope and restrictor sets, e.g., a binary existential quantifier might require that the intersection of the denotations of the two predicates is non-empty: $\exists^2\alpha(A, B)$ is true iff and only iff $|A| \cap |B| \neq \emptyset$. A binary quantificational *the* could require that the same intersection be a singleton set.

\forall^2 is the quantifier analogue of the strict conditional, $A \rightarrow B$, where $A \rightarrow B$ is true just in case all A-worlds are B-worlds.¹⁹ $\forall^2(A, B)$ is true just in case all A-individuals are B-individuals. But Ludovician counterfactuals are not strict conditionals, and strict conditionals do not really *need* systems of spheres for their interpretation. To find a quantifier analogue for the counterfactual we have to go a step further: we need a *variable* binary quantifier. Variable binary quantifiers demand that the specified relationship holds between the scope and restrictor sets, but *which* spheres it needs to hold throughout will vary, depending on which spheres contain individuals which satisfy the restrictor. Let’s use the notation $|A|_S$ to denote the set of individuals satisfying the formula A *which are members of sphere* S . Then, where α is an individual variable, and A and B are formulas:

$\forall^{2v}\alpha(A, B)$ is true in $M = (U, R, I)$ if and only if either

1. no individual in any sphere in R satisfies A .
2. some individual in some sphere $S \in R$ does satisfy A , and $|A|_S \subseteq |B|_S$.²⁰

Let’s look at some examples. Consider $\forall^{2v}x(Fx, Gx)$ and a model $M = (U, R, I)$. We are interested in the truth-value of the sentence in that model. There are several cases.

¹⁷ An A-world is a world where A is true.

¹⁸ Sider (2010) provides a useful introduction to binary quantifiers for philosophers.

¹⁹ $A \rightarrow B$ is equivalent to $\Box(A \rightarrow B)$ where \rightarrow is the standard material conditional.

²⁰ For many purposes we can simplify clause 2 so that it requires only that the *smallest* sphere in which A holds of some $u \in U$ is such that throughout it, if an object satisfies A , it satisfies B . But for the general case we use the definition above in order to handle cases where there is no smallest sphere where A holds of someone, perhaps because there are infinitely many people and infinitely many spheres where A holds. (Suppose, for example, that the human race will survive forever and everyone is ranked by birthdate (everyone is subordinate to their children and future descendants) and we take the domain to be the set of all humans who exist through out history and the future. If A holds for all individuals born in the future, there will be no smallest sphere where A holds.

Case 1: Vacuous Truth It might be that there are no F -individuals in any of the spheres in the model. Then the model makes $\forall^{2v}x(Fx, Gx)$ true by clause 1 and the sentence is *vacuously* true. (See the left-hand side of figure 1.) In such cases, the “opposite” generalisation, $\forall x(Fx, \neg Gx)$, is also true.²¹ An informal illustration might be *All werewolves are hungry* and its opposite *All werewolves are not hungry*. Both are true, but only because there are no werewolves.²²

Case 2: Non-Vacuous Truth It might be that there are F -individuals in one (or more) of the spheres in the model, and throughout at least one of those spheres with F -individuals, all the F -individuals are G -individuals. (This is consistent—importantly—with the existence of larger spheres in which there are F -individuals who are not G -individuals.) In this case the sentence is true in the model by clause 2 above, and its opposite, $\forall^{2v}x(Fx, \neg Gx)$, is false. (See the right-hand diagram in figure 1.) An informal illustration might be *All women are permitted to vote* on the assumption that there is a social sphere limited to white, adult, middle-class US citizens with no felony convictions and all women within *that sphere* are permitted to vote. Still, there could be larger spheres in which there are individuals who are women who are not permitted to vote, perhaps because they are incarcerated, or not citizens.²³

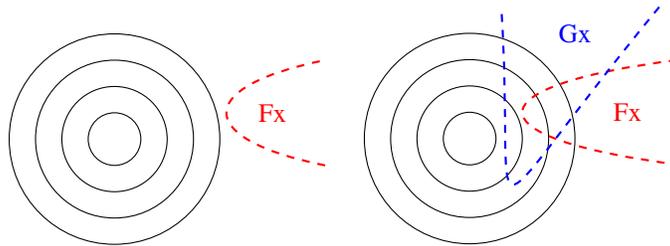


Figure 1: $\forall^{2v}(Fx, Gx)$: vacuous (left) and non-vacuous (right) truth

Case 3: If there are F -individuals in one or more spheres, but in every such sphere there is at least one F -individual who is *not* a G -individual, then $\forall^{2v}x(Fx, Gx)$ is false, as neither clause 1 nor clause 2 is met. $\forall^{2v}x(Fx, \neg Gx)$, may be true or false, depending on which of two subcases we are in. If one of the spheres containing an F -individual is such that all the F -individuals in it are $\neg G$, then $\forall^{2v}x(Fx, \neg Gx)$ is true. Call that case 3a. (See the left-hand diagram in figure 2.) But if all the spheres containing F -individuals are such that some F -individuals are G -individuals and some are $\neg G$ -individuals, then both $\forall^{2v}x(Fx, Gx)$ and $\forall^{2v}x(Fx, \neg Gx)$ will be false. Call that case 3b. (See the right-hand diagram in figure 2.) For an informal illustration of 3a, consider *All women are safe from harassment*. Plausibly there is a social sphere containing women

²¹In this “opposite” terminology I follow Lewis, who calls the counterfactual $A \square \rightarrow \neg B$ the opposite of $A \square \rightarrow B$.

²²In the informal illustrations we assume that ‘all’ is interpreted as a variable binary quantifier. I take it to be an interesting empirical question whether there are any such quantifiers in English.

²³Perhaps you think that means that the sentence *All women are permitted to vote* is *false*, but in that case you are refusing to read the sentence as containing a variable binary quantifier of the sort defined above.

such that every individual who is a woman is not safe from harrasment. Hence the sentence is false, and *All women are not safe from harrasment* is true. For an informal illustration of 3b, we might consider *All women like dogs*; on the assumption that the model contains no social class high enough that all the women in it like dogs, all the social spheres that contain women will contain some women who do not like dogs. Yet in this case the opposite generalisation is false as well: *All women don't like dogs* is false because no matter how high a social sphere we consider, some of the women in it like dogs and some do not.²⁴

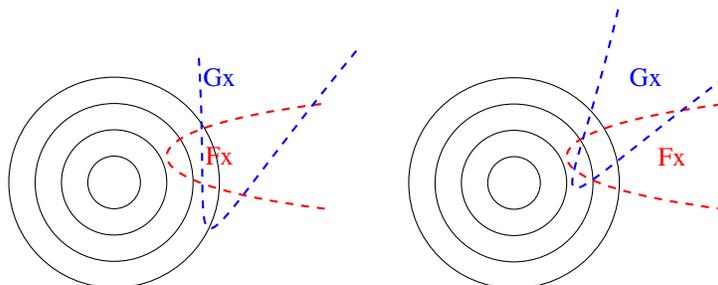


Figure 2: $\forall^{2v}x(Fx, Gx)$: Falsity with opposite true (left) and opposite false (right).

\forall^{2v} is just one of many variable binary quantifiers. There is also the quantificational counterpart of $\diamond \rightarrow$ —the *might*-counterfactual—as well as *the*, *most*, etc.

$\exists^{2v}\alpha(A, B)$ is true in a model (U, R, I) just in case some individual x in some sphere $S \in R$ is such that $x \in |A|_S \cap |B|_S$.

$Most^{2v}\alpha(A, B)$ is true in a model (U, R, I) just in case some individual in some sphere $S \in R$ does satisfy A , and $card(|A|_S \cap |B|_S) > card(|A|_S - |B|_S)$.²⁵

$The^{2v}\alpha(A, B)$ is true in a model (U, R, I) just in case there is a sphere $S \in R$ such that $card(|A|_S) = 1$ and $|A|_S \subseteq |B|_S$.

One application of a system of social spheres then, is to give satisfaction-conditions for variable binary quantifiers. Are there variable binary quantifiers in natural languages, like English? Perhaps. Consider the resident of Downton Abbey who notices a car outside and exclaims “A visitor has arrived!”, where they mean to express something which is true only if *someone socially important* has arrived, as opposed to a servant or tradesperson. Or consider what might be meant by “Most people have seen *Hamilton*,” “Most people have smart phones,” or “The discoverer of the Americas was Italian” (recalling that the Americas were discovered many times before Columbus.) Similarly “Peter suggested B” used in a context where several people suggested B, but Peter is the

²⁴In this case both $\exists^{2v}x(Fx, Gx)$ and $\exists^{2v}x(Fx, \neg Gx)$ will be true. These existentially quantified sentences are to the universally quantified ones as Lewis *might*-counterfactuals, written using $\diamond \rightarrow$, are to his *would*-counterfactuals.)

²⁵ $Card(X)$ denotes the cardinality of the set X .

only one of them who is a white male distinguished professor. It’s also possible that there are natural language sentences with these kinds of truth-conditions but without overt quantifiers—perhaps some generic claims made with bare plural constructions, such as “Birds fly” and “Deer ticks carry Lyme disease” (Cappelen and Dever, 2019, 129) have an underlying logical form which uses variable binary quantifiers.

5.3 Fallacies of Subordination

A third application of Lewis’ work is demonstrating counterfactual fallacies. There are argument forms containing counterfactuals which might have been expected to be truth-preserving—and which *are* truth-preserving when formulated with a strict conditional—but which Lewis’ model theory reveals to be invalid when formulated with $\Box \rightarrow$. Lewis identifies three: strengthening the antecedent, transitivity, and contraposition. In this section we will see that there are analogous fallacies for the variable binary universal quantifiers: strengthening the restrictor, quantifier transitivity, and quantifier contraposition.

5.3.1 Strengthening the Restrictor

To strengthen the antecedent of a conditional is to add conjuncts to its antecedent. If the original conditional was strict, then it entails the strengthened version:

$$\frac{A \rightarrow B}{(A \wedge C) \rightarrow B}$$

The premise requires every A -world to be a B -world, the conclusion—the stronger conditional—makes the weaker demand that every $A \wedge C$ -world is a B -world. The former is sufficient for the latter, and so the entailment holds.

Ludovican counterfactuals are different. They are *variable* binary operators, and which sets of worlds need to be B -worlds depends on where the antecedent is true. Since the antecedent varies between the two conditionals, space is made for the strengthened conditional to fail while the original continues to hold, making the following argument form invalid:

$$\frac{A \Box \rightarrow B}{(A \wedge C) \Box \rightarrow B}$$

Suppose there is a sphere with A -worlds, and all the A -worlds in that sphere are B -worlds. This makes the premise true. Suppose also that there are no $A \wedge C$ worlds in that sphere, though there are some in larger spheres. However, there is no sphere in which there are $A \wedge C$ -worlds in which it is also the case that all the $A \wedge C$ -worlds are B -worlds. This makes the conclusion false, though the premise remains true. For a more informal illustration, Lewis offers us:

$$\frac{\text{Kangaroos don't have tails } \Box \rightarrow \text{they fall over.}}{(\text{Kangaroos don't have tails } \wedge \text{they use crutches}) \Box \rightarrow \text{they fall over.}}$$

Lewis holds that the premise is true: in the closest spheres of worlds where kangaroos don't have tails, they indeed fall over. But in those spheres the kangaroos aren't using crutches—the no-tails-but-crutches worlds are overall less similar to ours than the no-tails-and-no-crutches worlds. And in such worlds (the $A \wedge C$ -worlds), kangaroos don't fall over, making the strengthened counterfactual false, and the argument invalid.

The quantificational analogue of strengthening the antecedent is *strengthening the restrictor*. If we do this with a strict (i.e. non-variable) binary universal quantifier, the result is valid:

$$\frac{\forall^2 \alpha(A, B)}{\forall^2 \alpha(A \wedge C, B)}$$

But with a variable quantifier, like the social ones we are interested in here, there are counterexamples:

$$\frac{\forall^{2v} \alpha(A, B)}{\forall^{2v} \alpha(A \wedge C, B)}$$

The restrictor of the premise can be satisfied by an individual in a smaller social sphere, while the stronger restrictor in the conclusion is only satisfied by an individual in a larger sphere. So long as the scope is satisfied throughout the smaller but not the larger, the premise will be true but the conclusion not, making the argument invalid.

Consider the following case in which subordination plays a role. 1905 is the date usually given for women's suffrage in Australia. So—reading the 'all' as a variable universal quantifier—the following quantified sentence became true in 1905: *All Australian women are permitted to vote*. But in fact it is only true if we are restricting our quantification to an inner, high status sphere of individuals, because *indigenous* suffrage lagged behind women's suffrage, and indigenous women were not permitted to vote in Australia (as a whole) until 1962. Hence in 1905 this quantifier with a strengthened restrictor was *not true*: *Indigenous Australian women are permitted to vote*. Here the strengthened predicate, 'Indigenous Australian woman' isn't satisfied by the individuals in the smaller sphere, and so we are triggered to look more broadly. We can formalise this as a counterexample to

$$\frac{\forall^{2v} x((\text{Australian}(x) \wedge \text{Woman}(x))(\text{AbleToVote}(x)))}{\forall^{2v} x((\text{Australian}(x) \wedge \text{Woman}(x) \wedge \text{Indigenous}(x))(\text{AbleToVote}(x)))}$$

Or more succinctly:

$$\frac{\forall^{2v} x((Ax \wedge Wx)(Vx))}{\forall^{2v} x((Ax \wedge Wx \wedge Ix)(Vx))}$$

Hence strengthening the restrictor is a fallacy with variable social quantifiers.

5.3.2 Quantifier Transitivity

The second fallacy Lewis discusses is *transitivity*:

$$\frac{A \Box \rightarrow B}{\frac{B \Box \rightarrow C}{A \Box \rightarrow C}}$$

He attributes the following counterexample to Stalnaker:

If J. Edgar Hoover had been born a Russian, he would have been a communist.
If J. Edgar Hoover had been a communist, he would have been a traitor.

If J. Edgar Hoover had been a Russian, he would have been a traitor.

The trick to formulating such counterexamples is to make the antecedent to the first a more remote possibility than the antecedent to the second. Then the consequent of the second can hold throughout a smaller sphere (making the second premise true) without holding throughout a larger sphere (making the conclusion false.)

The analogous form with variable quantifiers is:

$$\frac{\forall^{2v} \alpha(A, B)}{\frac{\forall^{2v} \alpha(B, C)}{\forall^{2v} \alpha(A, C)}}$$

If the quantifiers above were non-variable, the argument would be classically valid, in fact it would be the Aristotelian syllogistic form *Barbara*. But with variable quantifiers there is the possibility that the restrictor in the first premise (*A*) is satisfied only by individuals in a larger social sphere than the restrictor in the second (*B*). That allows the scope of the second (*C*) to hold throughout the smaller spheres, while there are individuals that satisfy *A* but not *C*—making the conclusion false. Here is a potential informal counterexample:²⁶

People who use sign-language gesticulate while speaking.
People who gesticulate while speaking should take a course on body language.

People who use sign-language should take a course on body language.

Someone uttering premise 2—say someone who coaches public speakers—might have in mind speakers using non-signed languages and perhaps not even be considering speakers of ASL and other signed languages. Users of signed languages gesticulate based on the meanings of gestures in those languages (they shouldn't do it differently because a course on body language suggested that it was “distracting” or made people think that they were “hiding something”.) But premise 1, by explicitly mentioning speakers of signed languages, widens the sphere of consideration to include such speakers. So both premises can be true, while the conclusion is false.

²⁶That is, it is a counterexample if the logical form of the sentences matches that of the formal argument above.

5.3.3 Quantifier Contraposition

Lewis' third fallacy is *contraposition*, which may refer to either of two forms:

$$\frac{A \Box \rightarrow B}{\neg B \Box \rightarrow \neg A} \qquad \frac{\neg B \Box \rightarrow \neg A}{A \Box \rightarrow B}$$

The analogous forms with variable quantifiers would be:

$$\frac{\forall^{2v} \alpha(A, B)}{\forall^{2v} \alpha(\neg B, \neg A)} \qquad \frac{\forall^{2v} \alpha(\neg B, \neg A)}{\forall^{2v} \alpha(A, B)}$$

Consider the version on the left. Had the quantifiers been strict, the argument would be valid. But the quantifiers are variable, and if the smallest sphere in which A is satisfied is strictly smaller than the smallest sphere in which $\neg B$ is satisfied, then the premise can be true because all A -individuals are B -individuals in the smaller sphere, while the conclusion fails because there are $\neg B$ -individuals which are A -individuals in the larger. Consider, for example, that since the 15th amendment to the US constitution we might want to say that the sentence *black people can vote* is true. Of course, there are “marginal” cases—you might not be able to vote if you are under 18, incarcerated for a felony, or if you are not a citizen. Now consider the “contrapositive” *people who can't vote are not black*. Assuming the predicate *can't vote* is not satisfied in smaller spheres, we look to larger spheres to find someone someone who satisfies predicate, and once we are quantifying over these spheres which contain persons of lower social rank, *of course* some of the people in them are black (some children are black, some people convicted of felonies are black etc.), and this makes *People who can't vote are not black* false. So the premise can be true in a model that makes the conclusion false, and the argument is invalid.

5.4 Pragmatics and Accommodation

The fourth application for Lewis' spheres is to model pragmatic phenomena that occur when counterfactuals are used in conversations. (Lewis, 1973, 66) On one use of the expression, *pragmatics* is the study of linguistic phenomena that require more than just the conventional linguistic meaning of expressions for their explanation. We might require context, a cooperative principle, or various other theories to account for the effect. Phenomena that count as pragmatic can include indexicals, quantifier restriction, presupposition, various kinds of implicature, insinuation, and metaphor. Though there are many pragmatic effects and theories of them, I will focus on one, with the aim of giving a clear illustration of the use of social ranking models to explain instances. That phenomenon is *accommodation*.

5.4.1 Accommodation and Comparative Similarity

Consider the following pair of counterfactuals, which Lewis takes from Quine:

- (4) a) If Caesar had been in command, he would have used the atomic bomb.

b) If Caesar had been in command, he would have used catapults.

Suppose that at most one of these counterfactuals is true. Do Lewis' truth-conditions for counterfactuals tell us which one it is? Lewis thinks not, and that is because his account makes reference to a system of spheres, which carries information about how possible worlds are ranked in respect of similarity. But similarity is vague and we have not yet pinned the relation down precisely enough to resolve the truth-value of either counterfactual: the truth-conditions make it explicit that this depends on the details of the system of spheres (which represents a precisification of the similarity relation) and this, Lewis thinks, may be determined by *context*.

In dealing with Quine's opposed counterfactuals about Caesar, context must of course be consulted somehow. [...] I could [...] call on context [...] to resolve part of the relation of comparative similarity in a way favorable to the truth of one counterfactual or the other. In one context, we may attach great importance to the similarities and differences in respect of Caesar's character and in respect of regularities concerning the knowledge of weapons common to commanders in Korea. In another context we may attach less importance to these similarities and differences, and more importance to similarities and differences in respect of Caesar's own knowledge of weapons. The first context resolves the vagueness of comparative similarity in such a way that some worlds with a modernized Caesar in command come out closer to our world than any with an unmodernised Caesar. It thereby makes the first counterfactual true. The second context resolves the vagueness in the opposite direction, making the second counterfactual true. Other contexts might resolve the vagueness in other ways. (Lewis, 1973, p. 67)

I want to stress some points about Lewis' view in this passage.

- comparative similarity is not an entirely determinate relation on possible worlds, but
- it may be precisified (made more determinate) by the conversational context
- some precisifications of the similarity relation will make 4a) true but not 4b). Others will make 4b) true but not 4a).

Following Lewis' later *Score-keeping in a Language Game* (1979) we can think of the conversational context as represented by something like a scoreboard—a sequence of values representing aspects of the context like the speaker, time, place, and possible world, a Stalnakerian Context Set, *C*, of assumptions that are in force at that stage in the conversation, as well as other values. Lewis' remarks suggest that there could be a position on the scoreboard for a value, $\$$, representing a system of spheres, a precisification of the relation of comparative similarity on possible worlds.

Scoreboards evolve to track the progress of conversations: the value for *speaker* may change when another participant takes a turn, and the value for *time* will continue to march on predictably. If a speaker asserts "I have to leave in 5 minutes for my bus" and all else is equal (no-one protests or says that the buses don't come on holidays) then

the claims that the person will leave in 5 minutes and is getting a bus will be added to C .

Outside of analytic philosophy, it is unusual for people to directly assert that the relation of comparative similarity on possible worlds is so-and-so (much as it is unusual for them to make direct assertions about social status.) But the value for $\$$ on a conversational scoreboard can still change in response to conversational moves. Suppose we are in a conversation about the US-Korean war in which the similarity relation is not yet determinate enough to decide which of 4a) and 4b) is true. On the precisification $\$_1$, 4a) is true, whereas on the precisification $\$_2$, 4b) is true. At time t_1 , someone asserts 4a): “If Caesar had been in command, he would have used the atomic bomb.” Just as the previous assertion presupposed that the buses are running, so the counterfactual assertion presupposes that the comparative similarity relation is one that makes 4a) true, say $\$_1$. All else being equal (no-one protests or says scornfully “oh, you think Caesar would have used a weapon he’d never even heard of, do you?”) then at t_1 , the scoreboard evolves to feature $\$_1$ as the value for comparative similarity on possible worlds.

This happens in accordance with Lewis’s rule of accommodation. Accommodation is a general phenomenon which applies to all kinds of utterances in conversations, not just to counterfactuals. When conversational participants say things which presuppose that certain other things are in place, then the missing presupposition becomes—all things being equal—presupposed in the context as a result of their utterance.

If at time t something is said that requires component s_n of conversational score to have a value in the range r if what is said is to be true, or otherwise acceptable; and if s_n does not have a range in the value r just before t ; and if such and such further conditions hold; then at t the score component s_n takes some value in the range r . (Lewis, 1979, p. 347)

If we like, we can build this scoreboard-sensitivity explicitly into our statement of the truth-conditions of counterfactuals,²⁷ giving us an account that reintroduces some of the explicit context-sensitivity of our simple unary \Box in section 5.1:²⁸

$A \Box \rightarrow B$ is true at a world w , given the function $\$$ on the scoreboard C , if and only if either

1. no A -world belongs to any sphere S in $\$_w$,²⁹ or
2. some sphere S in $\$_w$ does contain at least one A -world, and $A \rightarrow B$ holds at every world in S .

5.4.2 Accommodation and Social Rank

Social rank can be made more determinate through accommodation too. Consider the following opposed pairs of sentences:

²⁷I say, “if we like” because this isn’t inconsistent with the original statement of the truth-conditions given, which reflects a vagueness in the expression “truth-conditions”. In one sense, the truth-conditions give us a function from worlds and sets of spheres to truth-values, and in another they give us functions from worlds and scoreboards/contexts (which determine sets of spheres) to truth-values. It’s up to us which level we want to focus on, hence “if we like.”

²⁸Thanks to comments from Roy Cook for this suggestion.

- (5) (a) Anyone is permitted to call Dana the n-word
- (b) No-one is permitted to call Dana the n-world.
- (6) (a) Black people ought to call Rufus ‘master’.
- (b) Black people ought not to call Rufus ‘master’.

When Dana and Rufus are arguing at the beginning of *Kindred*, Rufus thinks the a-sentences are true and the b-sentences are false, and Dana thinks the a-sentences are false and the b-sentences true. Moreover this difference is a consequence of their disagreement about the relative statuses of black and white people. Rufus is presupposing a nested system of social spheres on which every white person is a member of a sphere that no black person is a member of. Call that R_1 . Relative to R_1 , the a-sentences are true and the b-sentences are false. Dana rejects R_1 .³⁰ Let’s suppose that the system of spheres Dana accepts is an egalitarian one, $R_2 = \{U\}$. (There is just one social sphere, and it contains everyone.) Relative to R_2 , the a-sentences are false and the b-sentences are true.

In *Kindred*, Dana and Rufus’ negotiation over status is doomed from the start; the society around them brutally reinforces acceptance of R_1 over R_2 . Dana gradually becomes aware that she has time-travelled to a more explicitly racist society, with social norms much closer to Rufus’ expectations than to hers. Just as she becomes aware that it is 1810s Maryland,³¹ so she also becomes aware of her new social environment and modifies her claims about social status in the light of *that*.

I sighed. “Kevin, I think we’d better demote me. In this time...” (61)

Dana’s sigh reflects, I suggest, not that she agrees that she is (objectively or morally) worth less than white people, but her recognition that she now finds herself in—and will have to get along in—a society that ranks her beneath them.

Still, the social ranking relation is frequently indeterminate.³² Moreover, conversations can proceed by assuming false things about it.³³ Suppose that instead of meeting in 1810s Maryland, Rufus and Dana first meet when both of them time-travel to a deserted beach in the 2110s, a time after the demise of the human race. There is no social hierarchy in the world around them to interfere with their negotiations over what to call each other, because—as they discover—they are the last humans left alive. Suppose Rufus utters:

³⁰In fact, given their relative ages at the time of this conversation (about 9 or 10 years old, vs 26 years old) she likely thinks she has a higher social status than Rufus. (p. 5, pp. 15–16)

³¹“Where are you from” asked the woman suddenly. “The way you talk, you not from round here.” The new subject caught me by surprise and I almost said Los Angeles. “New York,” I lied quietly. In 1815, California was nothing more than a distant Spanish colony—a colony this woman had probably never heard of. (p. 37)

³²Even in so rigidly stratified a society there is indeterminacy on certain questions of rank and permissions—especially for the new-comers Dana and Kevin. See e.g. the questions about where Dana ought to sleep. (Butler, 2004, pp. 89–102) I conjecture that social ranking relations are often more indeterminate around people who are new to the social grouping in some way, and that accommodation or resistance at such times can have particularly long-lasting effects.

³³That is, the conversation may presuppose things about the ranking relation which are not reflected in the social world in which the conversation takes place, much as an conversation might proceed based on false assumptions about the place of a record in the pop charts.

(7) You ought to call me ‘master’—after all, you want me to call you ‘black’.

I suggested earlier that such an assertion presupposes that Rufus is of higher rank than Dana; he is suggesting that their requests to be referred to a certain way are equivalent, and the best explanation of this is that he is assuming they were both just requests to be referred to appropriately. But *these* labels are appropriate only if Rufus is of higher rank than Dana. Lewis’ rule of accommodation tells us—all else being equal—that at the time of Rufus’ utterance the relevant scoreboard value gets updated to accommodate this presupposition. In this case that might mean it is updated to R_1 —Rufus’ racist social ranking.

But everything need not be equal. Suppose Dana protests:

(8) I’m not calling some teenage boy “Master!”

The protest is not enough to get R_2 added to the conversational scoreboard, but without the racist social environment placing its finger on the scale, it *is* sufficient to keep R_1 out. Moreover, Dana and Rufus’ negotiations over what each ought to be called can be understood as implicit negotiations over social rank.

As we did with counterfactuals, we can make this explicit in our statement of the truth-conditions for quantifiers:

$\forall^{2v} \alpha(A, B)$ is true relative to the set of social spheres R on a scoreboard C if and only if either

1. no individual in any sphere in R satisfies A .
2. some individual in some sphere $S \in R$ does satisfy A , and $|A|_S \subseteq |B|_S$.

6 Complications, Objections, and Further Questions

Isn’t it more complicated than that? It is natural to think, on a variety of grounds, that the models presented here are too simple. Though there is indeed ranking involved in subordination, the interactions between the phenomena of race, class, gender, age, disability, and ingroup are complicated. The subordination associated with different characteristics can compound. The ranking relation itself may be context-sensitive: sometimes people suggest that men are higher-ranked in the paid workplace, but subordinated in some child-care contexts. Perhaps American social hierarchies are different from those in the UK. Another complexity stems from the fact that some self-effacing people appear to do work to undo their own dominance and the subordination of others, and even to self-subordinate, work which is sometimes regarded as admirable (when a famous senior male does it) and sometimes as weak (when a junior woman does it.) A fourth complication is that the permissions that accrue with high status and erode with low include *conversational* permissions to do things like contribute to the conversation (“talk back”) and challenge or control the scoreboard.³⁴

I agree that subordination is complicated and complicated in ways that are not represented in the present paper. However, I do not think this is a good reason to

³⁴(Russell, 2019, §7.3.1)

abandon the project outlined here. My first reason for this is a methodological principle articulated in Williamson (2008):

Philosophy can never be reduced to mathematics. But we can often produce mathematical models of fragments of philosophy and, when we can, we should. No doubt the models usually involve wild idealizations. It is still progress if we can agree what consequences an idea has in one very simple case. Many ideas in philosophy do not withstand even that very elementary scrutiny, because the attempt to construct a non-trivial model reveals a hidden structural incoherence in the idea itself. By the same token, an idea that does not collapse in a toy model has at least something going for it. Once we have an unrealistic model, we can start worrying how to construct less unrealistic models. (Williamson, 2008, 186–7)

Progress in social philosophy—just as in other areas—is difficult enough, and humans are bad enough at it—that we should make use of all the available resources for analysing and clarifying ideas: logical, statistical, literary, empirical, linguistic etc. The topic of subordination adds additional difficulty because the issues involved are personal and important. Modeling of the kind attempted here is one kind of resource and the work of using it to nail down some of the details of subordination makes some consequences of central claims evident and forces us to articulate details that might otherwise have remained vague. It thereby seems to request of us justification for assumptions and draw attention to the absence of key explanations. Though attempts to construct simple mathematical models often raise new problems and challenges, that’s a welcome and expected consequence of their clarity, not a reason to abandon the approach.

My second reason not to be daunted by the complexity of real life subordination is that complexities often arise out of the interaction of one or more simpler mechanisms. In these cases, the best explanation of the complex phenomenon will employ multiple simpler mechanisms, and the work of articulating the simpler mechanisms will then indeed be a part of the work of explaining the complex phenomenon. In these situations we make progress when we get a simpler model on the table.³⁵

Does interpreting someone as using variable binary quantifiers let them off too easily? A different concern is that I am making a kind of strategic mistake. Interpreting someone as expressing a restricted universal quantifier (such as a variable binary universal quantifier) is often a way of interpreting them as expressing a generalisation that is true (e.g. *all women (who are privileged to a certain degree) are permitted to vote* rather than a more broadly interpreted *false* generalisation (e.g. *all women (unrestricted) are permitted to vote*.) But why should we be generous in interpreting people

³⁵One of the reasons that Lewis wants to model the ranking of possible worlds in terms of similarity is that he thinks that it is possible to capture some of the complicated effects this way, rather than merely consigning all the explanatory work to an unsystematic “it varies with context.”: Lewis rejects the suggestion that we cease our work with the thought that counterfactuals are strict conditionals whose domain varies with context because “that is not altogether wrong, but it is defeatist. It consigns to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the waste in that wastebasket.” (Lewis, 1973, p. 13)

who are only concerned about a privileged subset? Wouldn't it be better to interpret them as saying the *stronger* thing, so that we can point out that it is false?

This worry is based on two mistaken assumptions. To interpret someone as expressing a restricted quantifier—variable or otherwise—is not always to interpret them as expressing a weaker claim. Restricted *existential* quantifiers are stronger, not weaker than their unrestricted counterparts. And second, charity in interpreting one's opponents is not a strategic weakness, but rather a requirement of a successful countermove. If we give a less than generous interpretation of someone's view, they immediately have the option of saying "oh, that's not quite my view" and retreating to a weaker and more easily defended position. The most powerful criticisms in philosophy interpret their target claims in the most charitable way possible, and *then* show that even this best version of the position is still wrong. Interpreting someone as expressing the weaker, restricted generalisation may make their claim true, but it leaves them open to accusations of failing to take everyone into account, covering up this lack of consideration by using a (superficially) general construction, and engaging in fallacious argument on its basis.

A related issue is that variable binary quantifiers like \forall^{2v} are, in a way, a method of muddying the waters. They take the task of interpreting someone away from consideration of the explicit, surface-level syntax of the sentence and force us to look at context and subtler questions of interpretation. This makes variable binary quantifiers an important tool for deception—both intentional and unintentional deception of others, and deception of oneself. They allow us to express a claim that *looks a lot like* an unrestricted, non-variable universal quantifier, while ignoring inconvenient counterexamples, and then protest that we didn't mean to endorse the strong consequences of the more general claim if we are challenged. In general, I think, these kinds of mechanisms favour the powerful; recognising them and making them explicit is a step in the direction of exposing and countering them. So we should not hesitate to attribute the use of variable quantifiers where this makes sense.

Is subordination always morally bad? "Subordination" often sounds like a bad thing, and most of the examples of subordinative action and speech in this paper are morally bad—e.g. subordination on the basis of race, gender, disability etc. But it isn't clear that everything that fits the characterisation of subordination—ranking one person or group of people below others, depriving them of rights and legitimating discrimination against them—is morally bad. Sometimes it might be just, or at least morally neutral. For example, it might be just to rank one group of people below others on the basis of skills and experience in the workplace, on the basis of crimes committed against others, or to rank young children below their parents: it's not morally bad that 5 year olds are not permitted to drive, and that their parents are entitled to prevent them from driving. Similarly, some subordination might be temporary and merely in the interests of efficient social organisation, as when we grant the chair of a session the power to decide who has permission to speak during a talk. If we feel any hesitation about saying that the members of the audience are *subordinated* to the chair, it is only because 'subordination' might sound morally loaded. There is more than one reasonable regimentation of the terminology here, but my preferred one is to leave the definition of subordination as morally neutral. On this way of speaking, we say that

young children are subordinated to their parents, and the audience is subordinate to the chair of the session (or they should be) but it doesn't follow from this that anything is wrong. Still, we can note that 'subordination' has connotations of moral wrong because many important instances of subordination *are* examples of moral wrong-doing: e.g. subordination on the basis of race.

The social ranking relation isn't clearly enough defined. What exactly is the relation, information about which is encoded in the system of social spheres? I have been saying that it tells us whether one person or group ranks above another, but is this ranking a matter of power, or of moral status, or of money, or class, or something else entirely? Can we give necessary or sufficient conditions for 'x ranks at least as high as y'? And if not, have I failed to provide any useful explanation, because I have failed to define or be clear about one of the basic terms used in that explanation?

I decline to be clearer about one of the basic terms used in my explanation—at least here. The basic framework I have provided is compatible with many different substantive theories of the ranking relation, and I do not wish to tie it to any in particular. I introduce the ranking relation with examples from *Kindred* designed to enable the reader to get a sense of what I am talking about. The work of giving necessary and sufficient conditions for ranking is separable from the work undertaken here, which is, to use Stalnaker's memorable phrase "at a different level of abstraction."³⁶

Who is this work for? My last question came up frequently at the Workshop on Feminist Philosophy and Formal Logic for which this paper was originally conceived. Who is supposed to benefit from the kind of work presented here? The danger the question highlights, I think, is that ivory tower academics might use oppression as inspiration for academic work that is of interest only to a narrow and privileged academic audience. That privileged audience gains, while the oppression which inspired the work is untouched and its victims gain nothing. One might have two specific concerns in particular: one, which applies to philosophical work in general, is that philosophy is too slow, and too focused on the theoretical, to be useful for practical action, when it is practical action that oppression demands. There is no serious doubt that subordination on the basis of race, or gender, is wrong, and to put time and energy into more fine-grained theoretical work on the topic is to waste time and energy that could have been spent fighting oppression—not writing about it. And the second concern is that formal work is done in a language that is inaccessible to the oppressed, it is, as it were, a treatise on poverty in high church Latin—the choice of style is a barrier to it being of interest to all but a few.

I don't want to misrepresent my own confidence or overblow the value of this contribution, so what follows is really an articulation of what I hope work like this can do. The first part of my answer is that while I think the very *general* truth that subordination on the basis of race, gender, etc. is wrong is quite clear, it remains true that there is unclarity and dispute around more specific things, including whether particular actions and utterances are instances of such. This is no accident. Given that it is clear that subordination (on certain grounds) is wrong, it is important to agents to be able

³⁶(Stalnaker, 1999, p. 79) See also (Lewis, 1973, p. 91–2)

to plausibly deny that their actions are instances of subordination on those grounds. Agents become skilled at avoiding *transparently* unjustly subordinating. The resulting unclarity tends to favour the powerful, while the subordinated suffer harms of several different kinds: the core harm of being unjustly subordinated, the epistemic harm of feeling unclear or unsure about what has happened (see also Fricker (2006)), and the dialectical harm of being unable to articulate, explain and defend the truth to others.

It may seem bold to claim that the victims often feel unclear themselves about what has happened, but I think this is actually quite a common uncertainty. Consider a young woman from a traditionally subordinated group. She attends a small conference and finds it difficult to get along in certain ways: when she asks a question in a talk, the speaker laughs it off, but it seems to her that later in the session he treats the same question from a senior man very differently. When she joins a group of people talking during a break they seem unwelcoming, and when she asks a question, the group interprets her as having asked a more basic question than the one she asked. Essentially, she feels like she can't get them to treat her as having the professional status she has. Later, when she asks herself why, she can think of several competing explanations. It could be that the group just won't take someone of her gender/age/race seriously, so that these things are causally relevant to her subordination. But other possibilities suggest themselves as well, foremost of which is: perhaps they didn't listen to me because they don't think I'm *good enough*. But perhaps also she might have done something inadvertently that turned the group against her? Essentially she wonders, is it their racism/sexism, or is it something *she* did? She knows she was subordinated, but it isn't clear whether it was wrongful subordination. The victim may even be told by the group—falsely—that the subordination is earned and their fault.

In these circumstances, I think theoretical work can be of help to us, not just as disinterested academics, but as sufferers of subordination. If we can articulate the details of some of the mechanisms—whether that's gaslighting, mansplaining, or yes, subordinating quantifiers—and thereby come to recognise our own experiences as instances of a more general pattern—even when this means getting them clearly in focus *as* harms that were done to us—it can come as a relief. First, it can be a relief to be able to better understand the incident oneself and feel more epistemic confidence about it. Second, it can be a relief to discover that this is not merely a matter of one's own suspicion, but that it is supported by facts that are available to anyone. And, third it can be a relief to realise that these facts can be presented in such a way that even those who are motivated *not* to accept them will feel pressure to acknowledge them. For this reason, I think that theoretical work on subordination can help the subordinated.

The second concern was about style—perhaps formal logic is too inaccessible. This concern gets some (but not all) of its force from a social association between high status and formal work.³⁷ We tend not to associate formal logic with e.g. a disabled, black woman, and for that reason don't expect her to engage with it. Of course, similar things might be said for time-travel fiction. I hope that these expectations are partly false and that some unfairly socially subordinated people *just are* interested in logic, as they might be interested in sci-fi, Latin poetry, or cyclocross. But when I say that social as-

³⁷I don't mean to suggest anything about the demographics of actual logicians here, I mean this in the sense of Anderson (1995)'s *gender symbolism*: "which occurs when we represent nonhuman or inanimate phenomena as "masculine" or "feminine" and model them after gender ideals or stereotypes."

sociations are not the only problem, I mean to acknowledge that there is also a problem with access: neither exposure to logic, nor the prerequisite education, nor study-time nor support, nor faith in one's abilities and future are fairly distributed goods, but all and more are necessary to make a paper like this accessible. Two potential remedies suggest themselves: eliminate the social hierarchy—so that everyone who wants to is able to access the work—and/or we translate the work from the inaccessible format to some more accessible one. In the former case, I think this is absolutely possible, and in the latter I am not sure whether it is possible. But for now, I think of this paper as my best attempt to do the work itself, with the sharpest tools I have access to. What this question highlights is that the value of this paper presupposes the accomplishment of *other* difficult work: perhaps translation into more accessible terms, but certainly the destruction of unjust hierarchies.

References

- Adams, D. (1978). *The Hitchhikers Guide to the Galaxy*. BBC Radio 4, March, Fit the Third.
- Anderson, E. (1995). Feminist epistemology: An interpretation and a defence. *Hypatia*, 10.
- Butler, O. E. (2004). *Kindred*. Beacon Press, Boston.
- Butler, O. E. (2017). *Kindred: A Graphic Novel Adaptation*. Abrams ComicArts, New York. Adapted by Damian Duffy and John Jennings.
- Cappelen, H. and Dever, J. (2019). *Bad Language*. Oxford University Press, Oxford.
- Fricker, M. (2006). Powerlessness and social interpretation. *Episteme: A Journal of Social Epistemology*, 3.
- Langton, R. (1993). Speech acts and unspeakable acts. *Philosophy and Public Affairs*, 22(4):293–330.
- Lewis, D. (1973). *Counterfactuals*. Blackwell, Oxford.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359.
- Maitra, I. (2012). Subordinating speech. In *Speech and Harm: Controversies over Free Speech*, pages 94–120. Oxford University Press, Oxford.
- Russell, G. (2019). Subordinating speech and speaking up. *Oxford Studies in Philosophy of Language*, 1.
- Sider, T. (2010). *Logic for Philosophy*. Oxford University Press, USA.
- Stalnaker, R. (1999). *Context and Content*. Oxford University Press, Oxford.
- Williamson, T. (2008). *The Philosophy of Philosophy*. Blackwell, Oxford.