The Evolution of Intersectional Oppression

Cailin O'Connor, Liam K. Bright, and Justin Bruner

Intersectionality theory explores the special sorts of disadvantage that arise as the result of occupying multiple disadvantaged demographic categories. One significant methodological problem for the quantitative study of intersectionality is the difficulty of acquiring data sets large enough to produce significant results when one is looking for intersectional effects. For this reason, we argue, simulation methods may be particularly useful to this branch of theorizing because they can generate precise predictions and causal dependencies in a relatively cheap way, and can thus guide future empirical work. We illustrate this point through models which show that intersectional oppression can arise under conditions where social groups are disadvantaged in the emergence of bargaining norms. As we show, intersectional disadvantage can arise even when actors from all social categories are completely identical in terms of preferences and abilities. The conditions necessary to derive such disadvantage are relatively minimal. And when actors behave in ways that reflect stronger intersectional identities, the potential for disadvantage increases.

1 Introduction

Intersectionality theory is a branch of social theorising modeling the effects of occupying multiple demographic categories. Intersectional theorists are most often concerned to highlight aspects of social life that would be easily missed if one only studied the social effects of occupying one demographic category. For instance, it could be that being a woman affects the wage one is paid for a given task, and being black affects the wage one is paid; but being a black woman affects one's wage in a fashion that is not easily predictable from the aforementioned effects. Studies which focussed only on the effect of race or gender respectively on wages would then fail to detect the peculiar intersectional effect of race and gender on wage rates. To avoid missing out on such effects, intersectional theorists tend to advocate methodological approaches that pay equal attention to the consequences of occupying each demographic category under study, and especial attention to the results of the interaction of these demographic categories [Bright et al., 2016].

Intersectional theorists tend to be especially concerned with oppression, subordination, and disadvantage. In Crenshaw [1991]'s foundational text she discusses the way in which people who were both women and immigrants had especial difficulty, due to

occupying the intersection of those demographic categories, in accessing aid from domestic violence shelters. This is representative of the general pattern among intersectional theorists. As Collins and Chepp put it in their textbook entry: "[t]he first core idea of intersectional knowledge projects stresses that systems of power [...] cannot be understood in isolation from one another; instead, systems of power intersect and coproduce one another to result in unequal material realities and the distinctive social experiences that characterise them [Collins and Chepp, 2013, 60]. Intersectionality theory, then, is that branch of social theory concerned to study the effects of intersecting structures of oppression, where those structures oppress people qua members of demographic categories.

While the significance of such interacting structures of oppression may be granted in theory, studying these structures in practice has proven to be very difficult. One significant methodological problem for quantitative study of intersectionality is the difficulty of acquiring data sets large enough to produce significant results when one is looking for intersectional results. Suppose one wishes to study the effects of the interacting categories of disability, age, and race. Then for each racial group, age cohort, and gender under study, one must have a big enough population within their combined intersections so as to generate results. This can quickly increase the costs and difficulty of research, and for some populations may even make it impossible. Exploratory research in this field can thus be prohibitively expensive, but since most theoretical work in the area beforehand was focused on particular demographic categories there is relatively little theory to underlie more hypothesis driven research.

For this reason we think that simulation methods may be particularly useful for the intersectional theorist. Simulations allow theorists to develop precise causal dependencies in a relatively cheap way, and can guide future empirical work by providing well motivated predictions about the effects of interacting social structures for people occupying intersections of demographic categories. Along these lines, we present models which show that intersectional oppression can arise under conditions where social groups are disadvantaged in the cultural emergence of bargaining norms. As we demonstrate, intersectional disadvantage can arise even when actors from all social categories are completely identical in terms of preferences and abilities. And when actors behave in ways that reflect stronger intersectional identities, the potential for disadvantage increases.

Besides generating potential theoretical underpinnings for empirical work, our models make two other sorts of contributions. First, they generate 'how-possibly' stories for the emergence of special intersectional disadvantage. And, in particular, they show that relatively little is needed to generate it when humans develop bargaining norms and conventions. Second, as we elaborate in the discussion section, in applying our results to intersectionality theory it becomes clear that predictions from the theory are, in some ways, underspecified. We use our framework to make precise several notions of what it would mean to have truly intersectional disadvantage.

Our paper will proceed as follows. In section 2, we introduce and justify the modeling framework which we use in the rest of the paper. In section 3 we show how actors in multiple minority groups can be specially disadvantaged simply by dint of their double-

minority status. In section 4, we show how power differences between groups can lead to intersectional disadvantage, and how both power and minority status can interact in the emergence of bargaining. In 5 we discuss the implications of this work and the upshot for intersectionality theory.

2 The Evolution of Bargaining, Power, and The Cultural Red King Effect

When it comes to the cultural emergence of bargaining norms, recent work suggests that various features of social groups can impact the chances that such groups are advantaged or disadvantaged. In particular, it has been shown that minority status can lead to disadvantage under a number of conditions [Bruner, working paper, O'Connor and Bruner, working paper, O'Connor, working paper], and also that various forms of economic, or political power can likewise improve bargaining outcomes for groups [Bruner and O'Connor, 2015]. Here, we will use these results as a framework for exploring intersectional effects. In particular, we look at how intersectional identities can impact the emergence of discriminatory bargaining norms. In order to do this, we must first introduce this framework. We begin by describing the cultural Red King Effect, by which minority groups can be disadvantaged in the emergence of bargaining conventions, and then show how one form of economic power can likewise impact such conventions.

2.1 The Cultural Red King

The Red King effect occurs in biology when two mutualistic species co-evolve, but at different rates. Bergstrom and Lachmann [2003], using an evolutionary game theoretic framework, show that under some such conditions the slow evolving species can is more likely to end up getting more of the products of the mutualism.

Bruner [working paper] replicates this effect in a cultural context when a minority group interacts with a majority one. While minority types meet majority types commonly, as a result of their prevalence in the population, the reverse is not true. This means that the majority is very significant to the minority from a payoff standpoint, leading to a situation where minority members are more reactive to the majority group and learn more quickly how to interact with them. In cases of bargaining and division of resources, the minority can end up at a disadvantage as a result [Bruner, working paper, O'Connor and Bruner, working paper, O'Connor, working paper]. In the next section, as mentioned, we will investigate how this cultural version of the Red King effect can be especially disadvantageous to small intersectional populations. It will first be useful, however, to look at it in greater detail.

Evolutionary game theoretic models represent the evolution or emergence of strategic behavior in populations of actors, either biological or social. In particular, they have been of great use to philosophers in understanding the emergence of norms and conventions in human groups.[1]

[1]For examples see Skyrms [1994, 2004], Alexander and Skyrms [1999].

These models start with a *game*—a representation of strategic interaction that specifies *players* (who interacts), *strategies* (what strategic actions they can take), and *payoffs* (their outcomes from various sets of strategies). The games of particular interest for our purposes are those that approximately correspond to what economist Schelling [1980] referred to as 'mixed motive' games. In these games actors have some common interests in that they all prefer to coordinate their behavior, but their interests conflict with respect to how this coordination happens.

Bargaining is a classic example. Bargainers want to settle on a mutually agreeable division of a resource, but their preferences as to the details of this division differ. For this reason, and because norms surrounding the division of resources are deeply important to understanding distributive injustice, we focus on bargaining interactions here. Nash [1950] introduced what is now called the 'Nash demand game' to represent a broad class of bargaining scenarios. In this model, two players each demand some portion of a resource. If their demands are compatible, they each receive what they asked for. If their demands are incompatible in that they exceed the resource, actors are unable to agree on a division and receive a poor payoff often called a *disagreement point*.

Figure 1 shows what is called a payoff table for a mini-version of the Nash demand game.[2] Rows represent possible strategies for player 1 and columns for player 2. Entries to the table list payoffs for each player for each combination of strategies with player 1 listed first. The available strategies are to make a Low, Med, or High demand. For simplicity sake, we assume that the total resource is 10, that $L + H = 10$, and that $L \leq 5 \leq H$. This means that the demands always include 5, plus two compatible demands that perfectly divide the resource like 1 and 9, or 4.5 and 6.5. The disagreement point, which actors get for combinations of incompatible demands like Med v. High and High v. High is chosen to be 0.

		Player 2		
		Low	**Med**	**High**
Player 1	**Low**	L,L	L,5	L,H
	Med	5,L	5,5	0,0
	High	H,L	0,0	0,0

Figure 1: Payoff table for a general Nash demand mini-game.

One of the most important solution concepts in game theory, which will also be relevant to our social evolutionary approach, is that of a *Nash equilibrium*. Nash equilibria are sets of strategies in a game where no actor can unilaterally switch and improve their payoff. Because of this they are often stable in the sense that actors playing them will have no incentive to change. There are three Nash equilibria of this mini-game—Low v.

[2]The mini-game approach to bargaining, which looks at a small but representative subset of possible strategies for a more complex game, is commonly employed in models of the evolution of bargaining. (See, for example, Young [1993], Skyrms [1994].)

High, Med v. Med, or High v. Low.[3] At these pairings, if an actor demands more, they get the disagreement point, if they demand less, they get less.

Besides the game, the other feature of an evolutionary game theoretic model is the *dynamics*. These are rules that specify how the behaviors of some group of actors engaging in a strategic scenario will change over time. The most commonly used dynamics in evolutionary game theory are the *replicator dynamics* which suppose that successful strategies become more prevalent in a population while those that do poorly die out. It has been shown to bear deep formal relationships to explicit models of individual learning [Börgers and Sarin, 1997, Hopkins, 2002] and learning by cultural imitation [Björnerstedt et al., 1994, Weibull, 1997, Schlag, 1998] so we will employ it throughout the paper as a model of cultural evolution in populations developing conventions of bargaining.

At last we can present an example of the cultural Red King effect in a cultural evolutionary model. Suppose we have a population with two types of people—women and men, or white people and black people, or star bellied and plain bellied sneetches. Everyone in the population recognizes these social categories as relevant ones for choosing strategies and for picking imitative role models. Either type may be in the minority, or else the two types could be equally prevalent. Suppose further that individuals in this population regularly engage in bargaining scenarios that are well modeled by the Nash demand game. As time goes on, actors develop strategies for their bargaining behavior according to the replicator dynamics. We assume that they are repeating behavior that works well for them and for group members like them, and ceasing behaviors that generate low payoffs.[4]

If we evolve such models we find that separate conventions emerge for in-groups and out-groups. In particular, when members of either type interact with their in-group they come to do one of two things. Most commonly, they all eventually learn to demand Med, or to treat each other fairly. Sometimes they develop a fractious pattern of behavior where some actors demand High and some Low. Of more interest for our purposes are the conventions that emerge between the two types. These reflect the three Nash equilibria described above—either one type demands High and the other Low, or else they make Med demands of each other. Axtell et al. [2000], looking at similar models, describe the two sorts of outcomes where one type demands High of the other as discriminatory in the sense that actors treat out-group members differently than in-group members to the detriment of the out-group.

The three possible outcomes between groups occur with varying frequencies depend-

[3]To be completely clear, there are three pure strategy Nash equilibria. We will not address mixed strategy Nash equilibria because they are not relevant to our evolutionary results.

[4]The version of the replicator dynamics we employ is similar to the discreet time two population replicator dynamics, but where individuals interact with members of both populations and where one population may be smaller than the other. Let x and y represent the two populations so that $x + y = 1$ and $x \leq y$. Strategies for population x update according to $x_i' = x_i(\frac{f_i(x,y)}{\sum_{j=1}^{n} f_j(x,y)x_j})$ where x_i is the proportion of the x population playing strategy i, $f_i(x, y)$ is the fitness of type i in x given the population states of x and y, and $\sum_{j=1}^{n} f_j(x, y)x_j$ is the average population fitness for x given the states of x and y. Strategies for population y update according to the analogous dynamics.

ing on parameter settings of the models. The cultural Red King effect occurs when a type that is in the minority becomes more likely to end up being discriminated against. Under the replicator dynamics, this occurs when that discriminatory outcome has a larger *basin of attraction*, or set of starting population states that evolve to it. As described above, this happens because minority groups are more reactive to the majority population than the majority population is to them.

Figure 2 shows this effect. Results are proportions of 10k runs of simulations that end up at each of the three possible outcomes between types.[5] Let $L = 4$ (so that the three demands are 4, 5, and 6). Let p_1 be the proportion of the majority type in the population, and p_2 the minority, so that $p_1 + p_2 = 1$ and $p_2 \leq p_1$. As is evident from the figure, when $p_1 = .5$, meaning that the two types are equally prevalent, the most common outcome is the one where the two types make fair demands of each other. The two discriminatory outcomes are equally likely to emerge. As p_1 increases, three things happen. First, fair outcomes become increasingly unlikely. Second, the majority group becomes increasingly likely to end up demanding High. And third, the minority group becomes increasingly unlikely to demand High.

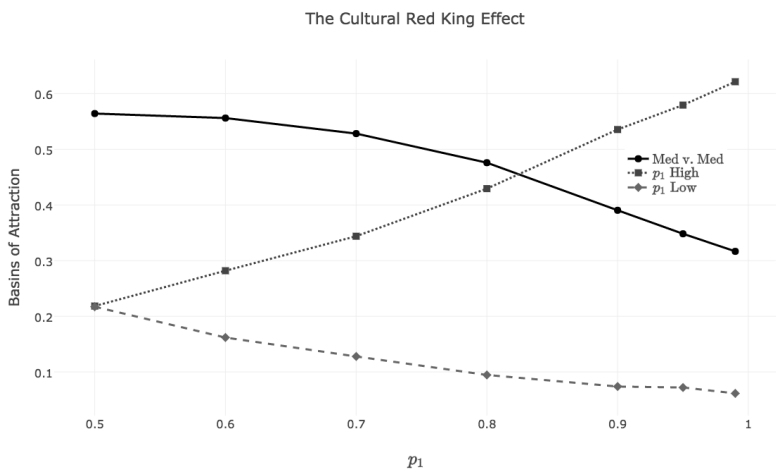

Figure 2: Basins of attraction for groups playing a Nash demand game as p_1 varies.

In other words, simply by dint of small numbers, members of a minority group can be disadvantaged by this effect. The cultural Red King effect is documented in much greater detail by Bruner [working paper], O'Connor and Bruner [working paper], O'Connor [working paper]. As these authors show, in some models a cultural Red Queen effect can also occur if, because of the details of their strategic situation, minority members quickly learn to make High demands and, as a result, gain an advantage from their reactivity. O'Connor [working paper], though, argues that if actors are risk averse or show in-group

[5]We use the discrete time replicator dynamics to generate these simulation results. For each trial the proportions of strategies in each type are randomly chosen and we evolve populations until they are converging to one of the three equilibria.

preference, common assumptions that hold of many real-world actors, the Red King is a much more powerful effect in the cultural realm.

2.2 Power and the Emergence of Bargaining

Nash [1953] used an axiomatic approach to argue that in rational choice scenarios of two player bargaining, a more powerful player can gain resources for themselves via a credible threat. Bruner and O'Connor [2015] explore not how actors can wield power to their benefit in a rational choice scenario, but how power differences between social groups can translate into an advantage with respect to emerging bargaining norms.[6]

In this paper, we will use a notion of power drawn from Nash's original work and employed by Bruner and O'Connor [2015] as well—that of asymmetric disagreement points. Suppose that when members of social groups bargain, and when bargaining breaks down, the fall back position for one group tends to be better than for the other. In economic work on the family, this assumption has been employed in models of household bargaining. Suppose that men tend to have a generally higher level of wealth, and better access to gainful employment, than women. When men and women bargain over household labor, then, the idea is that should the bargain break down (leading to household dissolution), men will tend to end up better off [Manser and Brown, 1980, McElroy and Horney, 1981]. Figure 3 shows a payoff table for a game integrating this assumption. This is the same game shown in figure 1, but where the disagreement point for player 1 is D, assuming that $D \geq 0$.

		Player 2		
		Low	**Med**	**High**
Player 1	**Low**	L,L	L,5	L,H
	Med	5,L	5,5	D,0
	High	H,L	D,0	D,0

Figure 3: A payoff table for a mini Nash demand game with different disagreement points.

Imagine two groups developing bargaining norms where one receives a higher disagreement point D. Although we will not show a figure outlining these results, in general, as D increases, the likelihood that the powerful group ends up at the equilibrium where they demand High also increases. The likelihood that they end up demanding Low or Med drops [Bruner and O'Connor, 2015]. In other words, their powerful situation translates into advantage vis a vis the emergence of bargaining norms. This is because groups with higher disagreement points are less incentivized to learn lower demands, since these demands do not yield outcomes that are significantly better than the disagreement point.

[6]These authors employ their results to explore how power differences in the academic hierarchy can translate to credit advantage in co-authored papers. Their results, though, are in fact quite general.

We will now employ these two sets of results to look at intersectional effects in populations where norms and conventions of bargaining emerge. In particular we will ask: How do intersectional identities impact results in these models? And, are small intersectional groups especially disadvantaged by the cultural Red King or power disparities?

3 Minority Status and Intersectional Oppression

We start with the cultural Red King and move through three sets of models where the importance of intersectional identity as a determinant of both behavior and cultural adaptation increases at each step. First, we compare two models where the second incorporates intersectional identity in a more full-blooded way than the first. As we will show, this leads to more serious disadvantage for small intersectional types. Then we present a model with even stronger intersectional assumptions. Although the results cannot be directly compared to the first two models (as we will elaborate), we see strong disadvantage for the smallest intersectional population. Across all three models, we find that the smallest intersectional population is specially disadvantaged solely by dint of their size.

3.1 Minimal Intersectionality

We start with a model that incorporates intersectional identities, but in a minimal way. Suppose that there are two dimensions of identity along which a population is split. For simplicity sake, call the prevalence of the majority and minority types along one dimension p_1 and p_2 and the prevalence of the types along the other dimension q_1 and q_2. The two dimensions of identity yield four resulting types. We assume that the two dimensions perfectly crosscut each other so that the proportions of the four types are $p_1 * q_1$, $p_1 * q_2$, $p_2 * q_1$, and $p_2 * q_2$. Figure 4 shows what these proportions might look like.

Because the models we will look at can get a bit complicated, we will use a silly example to help keep track of what is going on. (Later we will present a more realistic interpretation of the models.) Let's say our population is composed of Dr. Seuss's sneetches. The two relevant dimensions of social identity for bargaining behavior are belly stars—starred (p_1) or plain (p_2)—and color of the sneetch neck ruff—red (q_1) or gray (q_2). We assume that the star bellies and red neck ruffs will be the larger types. The four intersectional types are then red star bellies ($p_1 * q_1$), gray star bellies, ($p_1 * q_2$), red plain bellies ($p_2 * q_1$), and gray plain bellies, ($p_2 * q_2$). Figure 5 shows the four intersectional types in our fictional population.

Suppose that this population regularly engages in two sorts of bargaining scenarios, for each of which only one of their identities becomes salient. For example, one arena of bargaining could occur in the sneetch workplace over salary, benefits, or workload and, in this arena, belly stars could be particularly salient to the actors. Another arena of bargaining could occur in the sneetch marketplace over the cost of goods, and for this neck ruffs could be more salient.

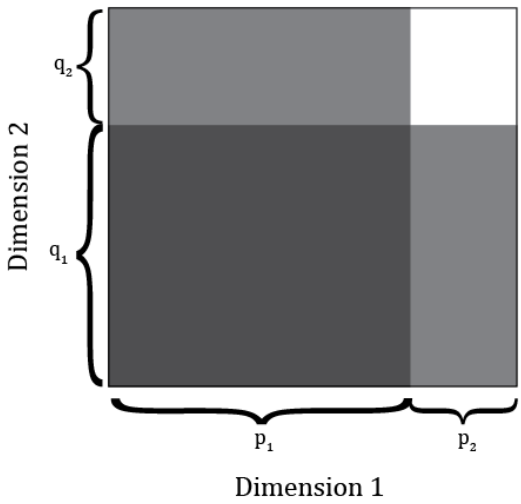

Figure 4: Diagram of proportional representation of intersectional types.

In this model, we expect that two separate processes will occur where bargaining norms conditioned on belly stars emerge in the workforce and norms conditioned on neck ruffs emerge in the marketplace. We do not expect these processes to impact each other, but in each arena we do expect the minority group to be disadvantaged in bargaining as a result of their size, i.e., for plain bellies to be more likely to get the short end of the stick in the office and gray ruffs in the market.

To keep things tractable, we will focus on an even smaller version of the Nash demand game, where actors may only demand High or Low, with payoffs otherwise identical to the game pictured in figure 1. There are four possible joint outcomes of the evolutionary processes in this model. First, both star belly and red ruff sneetches can demand High when interacting with plain and gray sneetches respectively, who demand Low in response. Second, star and gray can demand High. Third, plain and red sneetches can demand High. And last, plain and gray can demand High. Figure 6 shows the proportions with which these outcomes emerge in the model as the size of the majority types increases. We assume that the two majorities are the same size, or that $p_1 = q_1$.[7]

As is evident from the figure, as the sizes of the larger populations increase it becomes increasingly likely that star bellies and red ruffs demand High. It also becomes very unlikely that both plain bellies and gray ruffs demand high. And last, it becomes slightly less likely that the outcomes where stars and grays or plains and reds both demand High occur. These outcomes are a result of two cultural Red King effects, one for each of the evolving bargaining scenarios.

[7]This assumption allows us to easily display our results. We investigated models where p_1 and q_1 varied independently and generally the greater either one was, the greater the disadvantage for the small types.

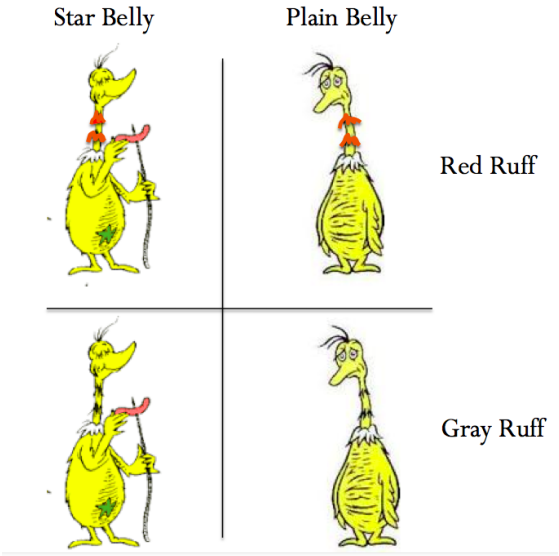

Figure 5: Four types of intersectional sneetches.

In this case it is most likely that members of the largest intersectional group end up making High demands in both arenas, and least likely that members of the smallest intersectional group do. This is arguably not truly an intersectional effect, though, as the outcomes are straightforwardly derived from two processes by which minority groups are disadvantaged. (In such a scenario, more robustly intersectional effects might occur if there are ways these processes combine to impact the lives of those involved. This could be the case if, for example, workplace bargaining determines what actors have to bargain over in the marketplace.)

3.2 Moderate Intersectionality

Now we consider a model with a stronger sort of intersectionality. Imagine a scenario as above where there are still two arenas in which bargaining occurs, and that each sort of identity is salient in one arena. Now assume, though, that actors only learn socially from those in their intersectional type. At work, for example, gray plain bellies do not assume that any plain belly sneetch is a good role model for them (even though belly stars are salient for interaction at work), but only adopt role models who also share their ruff type.

This small assumption greatly alters the cultural evolutionary process in these models. In particular, it means that the difference in learning speed between the largest and smallest types is much more significant in both arenas of interaction. This results in a much stronger Red King effect as is evident in figure 7. In the previous model, at the most extreme values of p_1 and q_1, the largest population ended up always demanding High about 45% and the smallest population 11% of the time. In this model these values

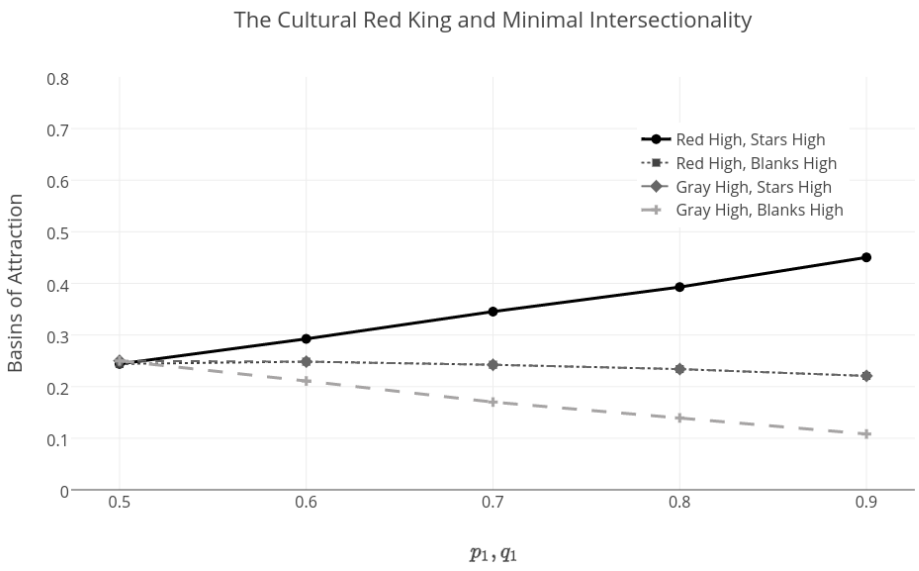

Figure 6: Proportions of four possible joint outcomes for minimally intersectional populations.

are now 76% and 1.5%.

In this case, the stronger intersectional assumptions in the model translate directly into a more significant disadvantage for the smallest intersectional group, and a stronger advantage for the largest group. In the next section, we look at a model with even stronger intersectional assumptions.

3.3 Strong Intersectionality

We now assume that there is just one arena of interaction for our intersectional sneetches, and that within this arena only intersectional identities are salient for interaction. In other words, the sneetches pay attention to intersectional identities in determining both 1) how to interact with bargaining partners and 2) which role models to choose.

The outcomes of this model are quite different from those presented in the last section. In- and out-group conventions now develop between *each* pair of the four possible intersectional types. This means that for each outcome of the model there are six types of out-group conventions, each of which may go one of two ways. Rather than report the proportions of runs that end up at each of these outcomes, we instead calculate the proportion of time each type spends demanding High of out-group members averaged over simulations.

Figure 8 shows results for this model. As in the last two models, we see that as the proportions of types grows more disparate, the disadvantage for the smallest type also becomes more significant, while the advantage for the largest type increases. In this

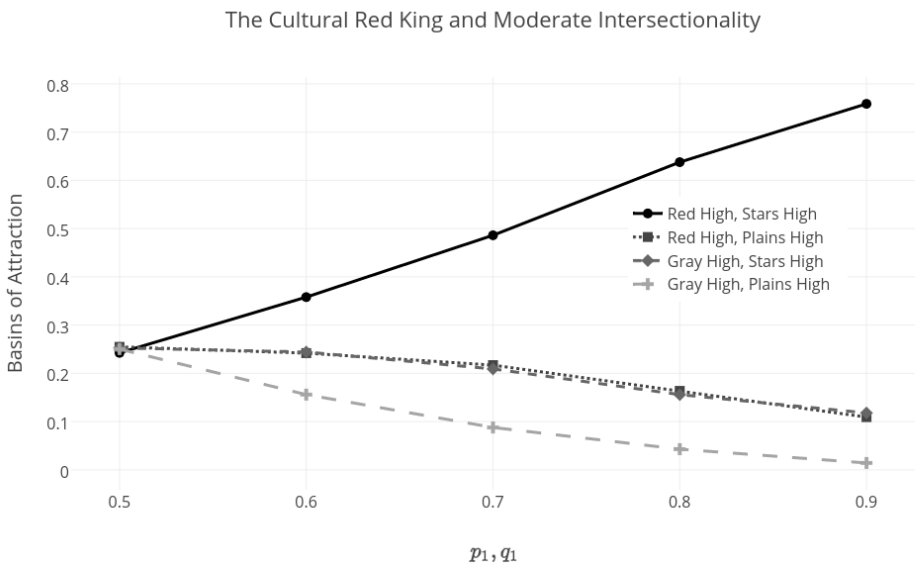

Figure 7: Proportions of four possible joint outcomes for moderately intersectional populations.

model, again, the effect is fairly dramatic. When $p_1 = q_1 = .9$ red star bellies demand High 90% of the time, compared to 9% of the time for gray plain bellies. This occurs because in a single arena, again, the size difference between the largest and smallest types is significant, leading to a strong cultural Red King.[8]

4 Power and Intersectionality

The models just presented tell us that under conditions where actors are members of two sorts of minority groups, and where actors take their intersectional identities seriously, members of small intersectional groups are at special risk of disadvantage as a result of dynamical effects.

In this section of the paper, we will now look at situations where there are power differences between groups. The goal here is twofold. First, we provide a robustness check on the sorts of intersectional effects we have been describing—similar patterns emerge even under different modeling choices. And second, we give examples of models that are perhaps a bit more realistic.

[8]Note that there is something unrealistic about this model, which is that the conventions developed need not be particularly consistent. For example, it could be the case that black women discriminate against white men, who discriminate against white women, who discriminate against black women. This sort of circularity does not occur in the two arena models.

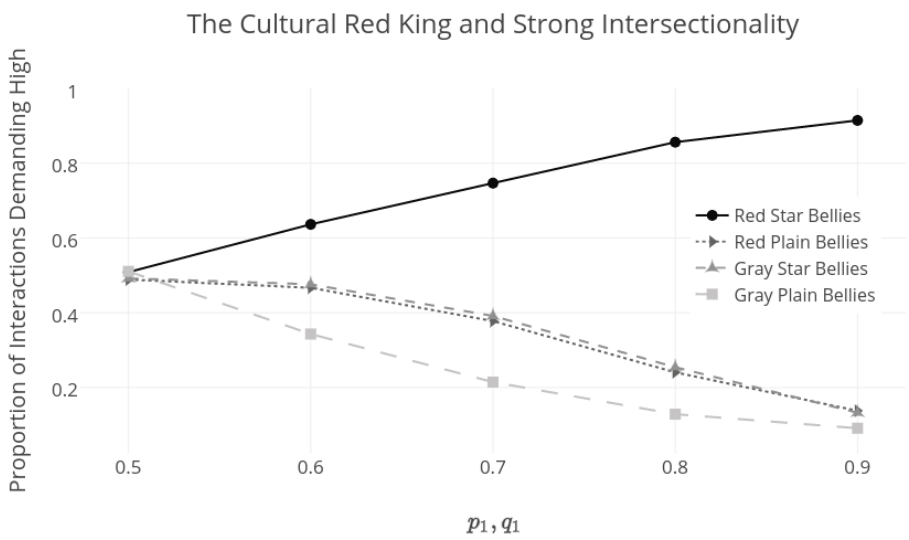

Figure 8: Average proportion of high demands for strongly intersectional populations.

4.1 Minority at Work, Powerless at Home

Let's first consider another two arena model like those discussed in sections 3.1 and 3.2. Forget the sneetches and consider a target situation where our two dimensions of identity are gender (man or woman) and race (white or black). Suppose now that the two arenas of interaction are the workplace, where race is more salient, and the home, where gender is more salient.

As before, let's assume that in the first arena of interaction—the workplace—black people tend to be in the minority so that the prevalence of white people, p_1, ranges from .5 to .99. And in the home, let's assume that men tend to have greater economic privilege, meaning their disagreement point for household bargaining, D, ranges from 0 to 3.9 while women's disagreement point is always 0. (We'll assume that $L = 4$ and $H = 6$ for this set of models, as in the previous ones.)

As in the previous set of models, we look at a case where actors display minimal intersectional identities when it comes to choosing role models (i.e., copy those of their race at work, and those of their gender in the home), and also a case where actors use role models only from their intersectional groups.

We find that the disadvantage generated by a cultural Red King effect at work, and by a power disadvantage at home can compound to create special intersectional disadvantage. Figure 9 shows results from the moderately intersectional models. We show a subsection of parameter values intended to capture the joint effects of power and minority status.[9] In particular, we show basins of attraction for the four possible

[9]We ran many other parameter values. In general, minority status and lower disagreement point led to increased disadvantage.

joint outcomes of the model (white people and men demand High, white people and women demand High, black people and men demand High, and black people and women demand High) as we increase the disadvantage for both black people with respect to proportional representation at work and women with respect to power at home. We use a bar graph here rather than a line graph because the x-axis represents two changing parameter values with no straightforward relationship to each other.

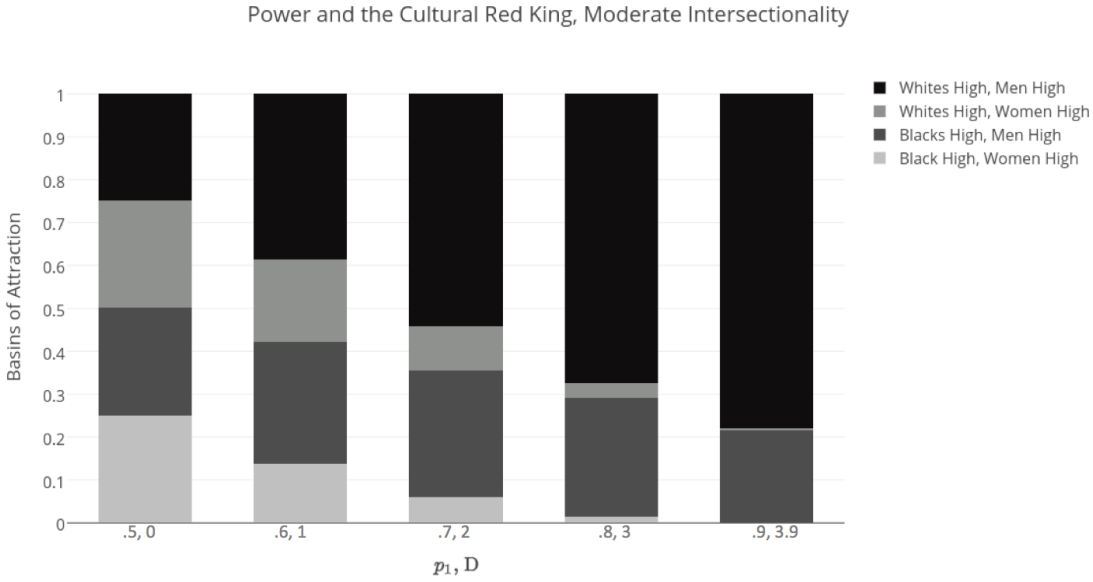

Figure 9: Basins of attraction for moderately intersectional populations with power and minority status.

As the figure makes clear, these variations lead to increasing disadvantage for women and black people, but especially for black women. At the most extreme values for disadvantage, they always demand Low in almost every simulation. White women are also strongly disadvantaged in these simulations because the effects of power are stronger than those of the cultural Red King.

As in the models solely relying on the cultural Red King, we also see a stronger intersectional disadvantage here when actors choose role models only from their intersectional types. For example, in figure 9, at the most extreme values white men always demand High in almost 80% of simulations. For the same parameter settings in the model with non-intersectional role models, white men always demand High in 63% of simulations. (We do not display these results for space purposes.)

4.2 Strong Intersectionality and Power

We also repeat the strong intersectionality model presented in section 3.3 but using power asymmetries, rather than size asymmetries, between groups. Remember that now

we simply have four intersectional groups that interact and develop bargaining norms. Actors pay attention to intersectional types both in choosing their strategies, and in choosing role models to imitate.

We now assume that the four types are equally prevalent in the population. But, because of the background conditions in their culture, there are power differences between them. We assume that white men have the most power so that in general when bargaining breaks down they will have a good fall back position and that black women are especially disempowered. We also assume that white women and black men occupy a middle ground, and for simplicity sake always give them the same amount of power. This is operationalized by making the disagreement point for white men always 3, for black women always 0, and for white women and black men varying it between 0 and 2.5.[10]

This might initially sound like a strange model. We start by assuming that black women are disadvantaged and then observe a disadvantage for them. But notice that we start with differences in power, and then look for subsequent disadvantage with respect to resources.

Figure 10 shows results from this model. The x-axis tracks the disagreement point for black men and white women. On the y-axis we see the proportion of time that each group demands High when meeting out-group members, averaged over all runs of simulation at that parameter value. When D for black men and white women is 0, all the intersectional groups (besides white men) are equally disadvantaged. As the disagreement points for the mid-power groups increases white men become slightly less likely to demand High in interaction, because now they are slightly more likely to end up disadvantaged with respect to these groups. Black women, though, become very unlikely to ever demand High. When D = 2.5, they only do so about 1% of the time.

Notice here that the trends related to disadvantage for black people and women generally do not capture how black women are disadvantaged in particular. When D is higher, in this scenario, while black men and white women do better on average, black women do worse. Part of the observation here is that in cases where intersectional types are truly all that matter, disjoint aspects of identity may have little power to predict disadvantage on their own.[11] Of course, in this model we freely vary the disagreement points of the four groups. If we assume that the disagreement point for black women is somehow a combination of those of black people and women, things will be different.

Before moving on to the discussion, we will pull out the sense in which these models display minimal conditions for the emergence of intersectional disadvantage, as described in the introduction. The generation of disadvantage here does not depend on representations of psychological features usually associated with discrimination. For example, the agents do not exhibit in-group preference and out-group bias, or implicit bias, or stereotype threat. We think these sorts of effects are important to full explanations of

[10] We looked at many other values for these parameters. In general, the higher the disagreement point for a group, the better for them, and vice versa.

[11] Note that in the strong intersectionality model in the last section, this was not the case because assumptions in the model meant that when the small intersectional group got smaller, this was a result of decreasing size of the two single dimensional groups making it up.

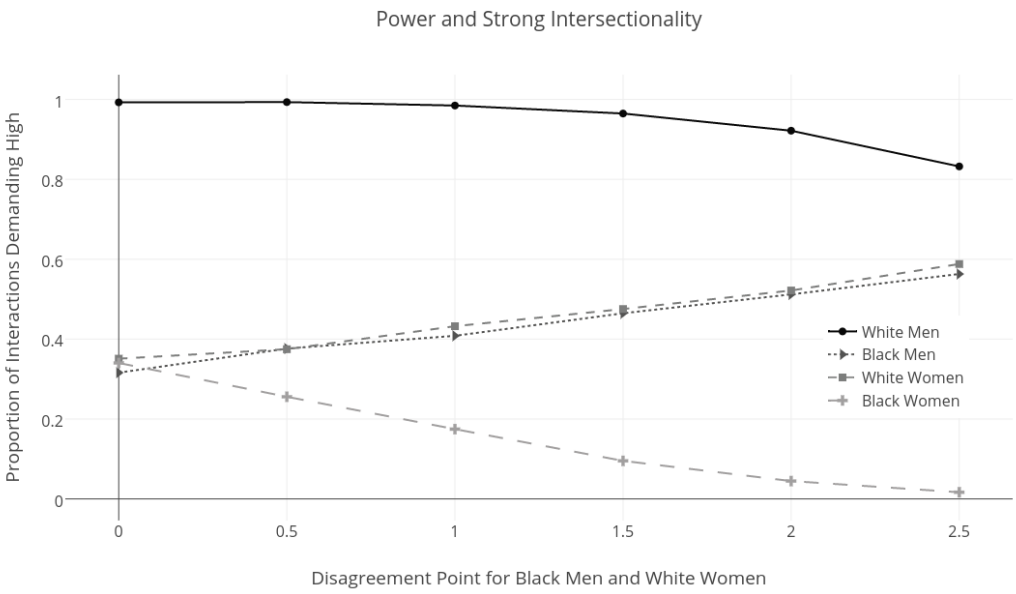

Figure 10: Proportion of interactions demanding High for four strongly intersectional groups.

real world oppression, but our models tell us that we can get it even without them. The preconditions necessary for oppression in our model are 1) intersecting minority demographics (or intersecting groups with power differentials) 2) actors who learn to behave in their best interest and 3) a strategic interaction that involves resource division. These minimal conditions tell us that intersectional disadvantage is easy to generate, underpinning claims by intersectional theorists that systems of power and oppression cannot be understood without understanding intersectional effects.

5 Discussion

The models we have presented show how intersectional groups can be disadvantaged by their joint identities via social dynamics. Our results demonstrate possible cultural evolutionary pathways for the emergence of intersectional oppression, and in doing so generate a set of testable hypotheses that can be further investigated through empirical means. To summarize, we found that the cultural Red King effect can produce special disadvantage for intersectional types, and that this disadvantage is worse when stronger assumptions of intersectional identity are made in the models. We also found that power and the cultural Red King can combine to produce intersectional disadvantage, and that power differences alone can lead to such disadvantage.

In order to fully understand how these results bear on intersectionality theory, though, we need to look more carefully at the theory itself. One of its core ideas is

that in some sense or another oppression isn't additive [Bright et al., 2016, 63]. Broadly speaking, this means that the combined effect of being oppressed along multiple axes is worse than one might expect if one were to naively sum the consequences of each form of oppression considered individually. We now investigate the sense in which the disadvantage we have observed in our models meets this criterion.

As it turns out, though, this is not a straightforward task. To accomplish it, one first needs sufficient theoretical background to generate a null hypothesis—by which we mean a theory of what it would look like for the oppressions in question to have a 'merely' additive effect. However, theories of oppression are not often formulated in sufficient detail to allow for the clear prediction, or even statement of, such a null hypothesis. Ironically enough, therefore, it may be useful for future work in intersectionality theory to precisify what a non-intersectional theory of oppression would look like.

We will present here two plausible precisifications of additive oppression to contrast non-additive effects with. The first criterion can be illustrated with a story. Suppose an employer encounters a job applicant who is black and a woman. Suppose further that they discriminate against black folk with probability P_b and women with P_w. If they are not being intersectional it may be as if they flip a coin (or are moved by what they had for breakfast that morning) to decide which is more salient to them, race or gender. They then focus on that aspect of the applicant in deciding whether to discriminate. In this case one would expect a level of discrimination somewhere in between P_b and P_w, because the combined chance of being discriminated against is a convex combination of the two chances of discrimination. With this understanding of non-intersectional oppression, we can say that there is non-additive intersectional oppression when the probability of being discriminated against given that you are a black woman is greater than both P_b and P_w.

A second, and quite different, null criterion is given by the following. In deciding whether to discriminate, the employer draws from first P_b and then P_w. If either proscribes discrimination, the employer discriminates. So if $P_b = .2$ and $P_w = .2$, the employer would discriminate against black women with probability .36. This is a second, stronger but plausible 'non-intersectional' level of discrimination.[12]

There may be other plausible criteria that one could use to determine whether some sort of discrimination or disadvantage for an intersectional group is non-additive. For the rest of this discussion we will use these two possibilities to analyze the results we have presented. For our models, we can calculate, given a set of parameter values, the probability that members of each group are discriminated against over all simulation outcomes. This probability is slightly different from the ones just described in motivating our criteria because it is averaged over a set of mutually exclusive outcomes. It does, however, capture something about the general chances of disadvantage for each group.

For all the models we have presented, except when groups sizes and disagreement points are equal, the most disadvantaged intersectional group suffers from non-additive

[12]Note that these two criteria will not capture null predictions for all cases of intersectional oppression. Consider empirical results showing that black women earn a lower wage than either white women or black men. Because these are not probabilistic results, neither of our senses of additive discrimination apply.

oppression in the first sense. I.e., the probability of being discriminated against if you are a gray ruffed, plain bellied sneetch, or a black woman, is always greater than the probabilities of discrimination for the two component identities. In general, as minority status becomes more dramatic, or as disagreement points for disadvantaged groups become lower, this difference in probability of discrimination for the intersectional group and the component identity groups also becomes more dramatic.

Some parameter settings of our models also show intersectional disadvantage in the second sense. Figure 11 shows various probabilities of discrimination for the model just mentioned. As we can see, for all values of majority size, $p_1 = q_1$, the probability of being discriminated against as a black woman, $P(D|Bl\&W)$, is greater than or equal to that for black people or women, $P(D|Bl)$ and $P(D|W)$, meaning it meets the first criterion as just described. When $p_1 = .9$, $P(D|Bl\&W) > P(D|Bl)\&P(D|W)$, or the probability of discrimination as a black woman is more than that expected if one met discrimination at the rate of both black people and women combined.

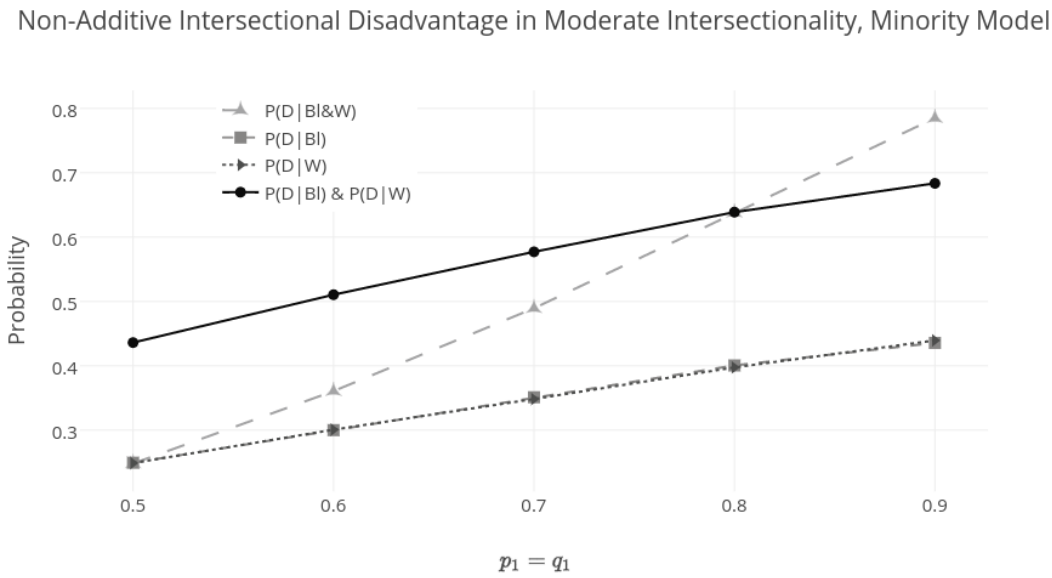

Figure 11: Probabilities of discrimination for different groups across parameter values.

Likewise, for the moderate intersectionality power models, when $D = 3.9$ and $p_1 = .9$, the probability of discrimination for women is .3 and for black people is .55. The probability of discrimination for both of these identities combined in the second sense would be .69, but the actual experienced probability of discrimination by black women is .75, meaning there is again special intersectional disadvantage in both of the sorts of ways we identify.

The take-away from this discussion is two-fold. First, our attempt to specifically quantify intersectional disadvantage helped us precisify just what it might mean for

there to be no intersectional disadvantage. The fact that we have generated two possible criteria for this null effect shows how current theories of intersectional disadvantage are, at least to some degree, underdetermined as to what they predict. Second, we find that the sorts of effects we have been outlining can generate non-additive intersectional disadvantage in either of the two ways we describe. On the less demanding criterion, all the models we consider generate such a disadvantage.

Intersectionality theory is at heart an empirical and engaged theory. Empirical, in the sense that it is a theory of the material conditions and life experiences of oppressed people in the social world as-is. Engaged, in that intersectionality theorists typically want not just to study but also to identify strategies for ameliorating structures of oppression. We hope we have shown that, despite their apparent remove from the lifeworld and its concerns, the modeling tools deployed in this paper can generate insights that may be of use to intersectional theorists in their core projects.

Bibliography

Jason Alexander and Brian Skyrms. Bargaining with neighbors: Is justice contagious? *The Journal of philosophy*, pages 588–598, 1999.

Robert Axtell, Joshua M Epstein, and H Peyton Young. The emergence of classes in a multi-agent bargaining model, 2000.

Carl T Bergstrom and Michael Lachmann. The red king effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences*, 100 (2):593–598, 2003.

Jonas Björnerstedt, Jörgen W Weibull, et al. Nash equilibrium and evolution by imitation. Technical report, 1994.

Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997.

Liam Kofi Bright, Daniel Malinsky, and Morgan Thompson. Causally interpreting intersectionality theory. *Philosophy of Science*, 83(1):60–81, 2016.

Justin P Bruner. Minority disadvantage in population games. working paper.

Justin P Bruner and Cailin O'Connor. Power, bargaining, and collaboration. In Conor Mayo-Wilson Thomas Boyer and Michael Weisberg, editors, *Scientific Collaboration and Collective Knowledge*. Oxford University Press, 2015.

Patricia H. Collins and Valerie Chepp. Intersectionality. In Johanna Kantola Georgina Waylen, Karen Celis and S. Laurel Weldon, editors, *The Oxford Handbook of Gender and Politics*, chapter 2, pages 57–87. Oxford University Press, 2013.

Kimberle Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6):1241–1299, 1991.

Ed Hopkins. Two competing models of how people learn in games. *Econometrica*, 70 (6):2141–2166, 2002.

Marilyn Manser and Murray Brown. Marriage and household decision-making: A bargaining analysis. *International economic review*, pages 31–44, 1980.

Marjorie B McElroy and Mary Jean Horney. Nash-bargained household decisions: Toward a generalization of the theory of demand. *International economic review*, pages 333–349, 1981.

John Nash. Two-person cooperative games. *Econometrica: Journal of the Econometric Society*, pages 128–140, 1953.

John F Nash. The bargaining problem. *Econometrica: Journal of the Econometric Society*, pages 155–162, 1950.

Cailin O'Connor. The cultural red king effect. working paper. Working Paper.

Cailin O'Connor and Justin P Bruner. Dynamics and diversity in epistemic communities. working paper. Working Paper.

Thomas C Schelling. *The strategy of conflict*. Harvard university press, 1980.

Karl H Schlag. Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, 78(1):130–156, 1998.

Brian Skyrms. Sex and justice. *The Journal of philosophy*, pages 305–320, 1994.

Brian Skyrms. *The stag hunt and the evolution of social structure*. Cambridge University Press, 2004.

Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.

H Peyton Young. An evolutionary model of bargaining. *Journal of Economic Theory*, 59(1):145–168, 1993.