

The Dark Side of Morality: Group Polarization and Moral-Belief Formation

Marcus Arvan
University of Tampa
marvan@ut.edu

Most of us are accustomed to thinking of morality in a positive light. Morality, we say, is a matter of “doing good” and treating ourselves and each other “rightly.” However, moral beliefs and discourse also plausibly play a role in *group polarization*, the tendency of social groups to divide into progressively more extreme factions, each of which regards other groups to be “wrong.”¹ Group polarization often occurs along moral lines², and is known to have many disturbing effects, increasing racial prejudice among the already moderately prejudiced³, leading group decisions to be more selfish, competitive, less trusting, and less altruistic than individual decisions⁴, eroding public trust⁵, leading juries to impose more severe punishments in trials⁶, generating more extreme political decisions⁷, and contributing to war, genocide, and other violent behavior.⁸

This paper argues that three empirically-supported theories of group polarization predict that polarization is likely caused in substantial part by a conception of morality that I call the *Discovery Model*—a model which holds moral truths exist to be discovered through moral intuition, moral reasoning, or some other process. §1 of this paper clarifies the Discovery Model, showing how it is ubiquitous in everyday life and moral philosophy, cohering as well with empirical research on how people ordinarily form moral beliefs. §2 then argues that three dominant

¹ See Isenberg (1986) and Pruitt (1971).

² See Haidt (2012).

³ Myers and Bishop (1970).

⁴ Luhan, Kocher, and Sutter (2009).

⁵ Rapp (2016).

⁶ Bray and Noble (1978)

⁷ Walker and Main (1973)

⁸ Newman (2002), Sunstein (2002).

empirical theories of group polarization—Social Comparison Theory⁹, Informational Influence Theory¹⁰, and Self-Categorization Theory¹¹—all predict that the Discovery Model likely plays a significant role in causing polarization. Finally, §3 argues that there are converse theoretical reasons to believe that an alternative *Negotiation Model* of morality—one according to which most moral truths are instead created by interpersonal negotiation—would likely mitigate polarization and perhaps even foster its opposite.

As a point of clarification, this paper’s aims are modest.¹² First, it aims neither to establish that the Discovery Model causes polarization, nor that the Negotiation Model is psychologically realistic or would reduce polarization. Because these are complex issues—and different possible explanations of polarization exist—this paper instead aims to provide strong *theoretical* grounds for investigating these matters further in future research. Second, this paper does not aim to settle a variety of important philosophical questions, such as whether group polarization is morally undesirable, or whether the Negotiation Model should be favored over the Discovery Model on meta- or normative-ethical grounds. Although I have defended Negotiation Model elsewhere on both grounds¹³, this paper cannot settle these wide-ranging issues. Instead, it once again aims to provide new theoretical reasons—grounded in the psychology of group polarization—to investigate these and other related questions in future research.

1 The Discovery Model of Morality

⁹ Bray and Noble (1978), Mackie (1986).

¹⁰ Stoner (1961); Myers and Arenson (1972); Hinsz and Davis (1984).

¹¹ Abrams et al (1990).

¹² I thank two anonymous reviewers for inviting me to rethink and foreground my argument’s scope.

¹³ In [book title redacted], I argue that seven meta-ethical principles of theory-selection support a new normative ethical theory in line with the Negotiation Model.

Some philosophers¹⁴ and everyday laypeople purport to be moral skeptics, alleging that there are no moral facts. Nevertheless, just about everyone plausibly forms moral beliefs in the course of everyday life. We regularly speak of people doing “right”, “wrong”, “good”, and “bad.” We also tend to do so in accordance with a particular model of morality: a Discovery Model according to which moral truths *exist to be discovered* through intuition, moral reasoning, or some other cognitive or affective process. The Discovery Model, as I propose we understand it, does not hold that we come to believe moral propositions passively or unreflectively. It is instead the conjunction of the following meta-ethical and psychological claims:

- **The discovery model of meta-ethics:** there are preexisting truths about moral issues (e.g. truths about right, wrong, good, bad) that can in principle be ascertained by individuals “unilaterally”, via their own use of intuition, philosophical argument, or some other cognitive or affective process.
- **The discovery model of moral-belief formation:** individuals who tacitly or explicitly endorse the discovery model of meta-ethics will tend to *form moral beliefs* through intuition, argument, or other such process, typically believing in at least some cases that they have discovered moral truths that other people should believe as well.

We can see just how ubiquitous the Discovery Model is by examining everyday discourse, the history and present of moral philosophy, and finally, social-psychological research.

Consider first everyday moral practice. When it comes to applied moral issues—such as whether abortion is morally permissible—people commonly presuppose that there are *moral facts* to be discovered through intuition, reasoning, or some other cognitive or affective process, typically forming *beliefs* about their “moral discoveries.” For example, individuals who believe

¹⁴ See e.g. Joyce (2007, 2015) and Mackie (1977).

that abortion is morally wrong often claim to base this “discovery” on the intuition or some argument that human beings have a *moral right to life*—arguing that since fetuses are human beings, we should all believe that abortion is wrong.¹⁵ Conversely, people who believe abortion is morally permissible typically base their belief on different intuitive or argumentative “discoveries”, such as that fetuses do not have a moral right to life at certain stages of development¹⁶, or alternatively, that a fetal right to life is not a right to depend on a mother’s body.¹⁷ Similarly, consider the Israeli-Palestinian conflict. Here too, both sides appear to treat moral truths as discoverable through intuition or argument. Whereas pro-Palestinian voices defend the moral “discovery” Israel has *wrongly* occupied Palestinian lands, pro-Israelis typically defend the opposite “discovery”—that Israelis have *rightly* occupied Palestine, claiming that settlers are simply “living on land that Israel has liberated.”¹⁸ In each of these cases, we see the Discovery Model’s meta-ethical and psychological components clearly exemplified. And these are not isolated cases. Indeed, the Discovery Model appears presupposed by every major world religion—with Judaism holding that we can discover moral truths through the Ten Commandments; Christianity that we can discover moral truths via Christ; Islam that moral truth is to be discovered through the Quran and/or Shari’a Law; Buddhism that moral truth is to be found in the Noble Eightfold Path; and so on.

The Discovery Model similarly pervades the history and present of moral philosophy. In contemporary meta-ethics, the notion that moral facts exist to be discovered is central to many (though not all) theories of moral semantics and moral epistemology. Indeed, although there are non-cognitivist interpretations of moral language, the dominant semantic view of moral language

¹⁵ See e.g. NRLC (2016), Pro-Life Perspective (2016).

¹⁶ See e.g. Arthur (2001).

¹⁷ See e.g. Liberty Women’s Health (2016)

¹⁸ See Black, Wedeman, and Mullen (2015) for a brief overview.

is cognitivist: the view that sentences of the form, “X is morally wrong”, are true just in case it is a *fact* that X is morally wrong.¹⁹ In addition, many moral realists argue that cognitivism is not only true, but that moral facts are mind-independent features of the world *discoverable* by us.²⁰ Further, consider a few dominant theories of moral epistemology: moral intuitionism, reflective equilibrium, and constitutivism. Intuitionists hold that moral truths can be discovered through a kind of moral perception analogous to sense-perception.²¹ In contrast, reflective equilibrium treats moral facts as things we can discover by drawing our moral beliefs into greater coherence, holding that we should form new moral beliefs as a result of reflective argumentation.²² Finally, constitutivists argue that moral truths can be discovered—and moral beliefs formed—by reference to constitutive features of agency.²³

Now consider normative ethical theory. Here too we see the ubiquity of the Discovery Model. Act-utilitarians hold that an action is right if and only if the act maximizes utility—facts that can in principle be discovered. Kantians hold that an action is permissible if and only if its maxim can be willed as a universal law²⁴, respecting the humanity of oneself and others²⁵—facts that once again can be discovered.²⁶ Aristotelian virtue ethicists hold moral virtues are beneficial character traits which we can discover to be necessary for living well.²⁷ And so on. Many other influential approaches to normative ethics—contractualism, Rossian pluralism, moral

¹⁹ See van Roojen (2015) for an overview.

²⁰ See e.g. Shafer-Landau (2003), Brink (1989), Dancy (1986), Finlay (2007).

²¹ See e.g. Audi (2015) as well as Stratton-Lake (2014): introduction, for an overview of intuitionism’s resurgence.

²² See Daniels (2013), Timmons (2007): 27-31, Vaughn (2009): 46-7, and Barcalow (2007): 14-5 for brief summaries of how pervasively reflective equilibrium is used to evaluate moral theories and arguments.

²³ See e.g. Kant (1785, 1797), Korsgaard (2008, 2009), and Katsafanas (2011).

²⁴ Kant (1785): 4:421.

²⁵ Ibid: 4:429.

²⁶ Ibid: 4:422-3, 4:429-31.

²⁷ See *Nicomachean Ethics*. Also see Hursthouse (1999): ch. 1.

particularism, etc.—similarly hold that moral truths can be discovered by intuition, argument, or some other cognitive or affective process.²⁸

The same is true in applied ethics. In the applied ethical literature on abortion, some argue that abortion can be discovered to be wrong because it violates the moral rights of the fetus²⁹; others argue that abortion can be discovered to be morally permissible at certain stages of fetal development due to the fetus not being a person³⁰; others still argue that abortion is permissible because a fetal right to life is not a right to depend on a mother's body.³¹ And so on. Once again, this is not an isolated case. The applied ethics literature is replete with works arguing that we can discover truths about applied ethical issues—about the ethics of torture, gun control, warfare, treatment of animals, etc.—through intuition, argument, or other cognitive or affective processes.

Finally, the Discovery Model's psychological account of moral-belief formation coheres with the dominant empirical model of moral-belief formation in social psychology: the Social Intuition Model (SIM) which holds that moral beliefs are the result of sudden flashes of affectively laden intuitions, with moral reasoning largely serving a subservient role to justify one's beliefs *ex post facto*.³²

2 The Discovery Model and Group Polarization

Group polarization, once again, is the tendency of social groups to divide into progressively more extreme factions, each of which regards other groups to be “wrong.” There are two leading empirical theories of the causes of polarization: Social Comparison Theory and Informational Influence Theory. According to Social Comparison Theory, polarization results from people

²⁸ See e.g. Scanlon (1998): 4, 191; Ross (1930); Dancy (2013); and Parfit (2011): §49.

²⁹ See e.g. Pojman (1998) and Marquis (2007).

³⁰ See e.g. Warren (1973).

³¹ Thomson (1976).

³² Haidt (2001).

seeking to “fit in” with those around them, aiming to impress members of their group by endorsing progressively extreme views.³³ Informational Influence Theory holds that, in addition, polarization results from people hearing new arguments and information in support of their position—processes that make group members more receptive to progressively more extreme views.³⁴ Importantly, these two theories are not mutually exclusive, nor are they “mere theories.” Because both theories have significant empirical support³⁵, it is more correct to say that they have identified *two primary causal mechanisms* of polarization. Finally, a third theory, Self-Categorization Theory, also has some empirical support.³⁶ This theory holds that polarization results from individuals identifying with the prototypical view of their “in-group”—a group which then casts views of “out-groups” as threatening, causing the views of the in-group to shift even further away from those of the outgroup as a kind of defense mechanism.³⁷ Allow me now to briefly explain each of these three theories in more detail.

According to Social Comparison Theory, individuals in groups have a psychological tendency to want to gain acceptance and be perceived favorably by other members of their group. This desire for acceptance causes individuals in the group to adopt ever-so-slightly more extreme views than those already typical in the group, so as to “impress” and “prove” themselves to other members of the group³⁸—something which often takes the form of *moral grandstanding*.³⁹ Group polarization then occurs when many individuals in the group do more or less the same thing, progressively adopting more extreme views to impress each other, causing the representative

³³ Bray and Noble (1978), Mackie (1986).

³⁴ Stoner (1961); Myers and Arenson (1972); Hinz and Davis (1984).

³⁵ For major confirming evidence of Social Comparison Theory, see Bray and Noble (1978), Myers and Bishop (1970), and Luhan et al (2009). For major confirming evidence of Informational Influence Theory, see Stoner (1961), Myers and Arenson (1972), Kaplan (1977), and Hinz and Davis (1984).

³⁶ See Hogg et al (1990) and McGarty et al (1992).

³⁷ Abrams et al (1990).

³⁸ Bray and Noble (1978), Myers and Bishop (1970), and Luhan et al (2009).

³⁹ See Tosi & Warnke (2016).

beliefs of the entire group to become progressively more extreme. Finally, and importantly, studies indicate that this phenomenon is even more likely to occur with respect to “judgmental issues”, such as moral or political matters.⁴⁰ For instance, a recent study on Twitter regarding the shooting of an abortion doctor indicated that like-minded individuals on both sides of the issue tend to group together, reinforcing and progressively polarizing pro-life and pro-choice views.⁴¹

Informational Influence Theory supplements this account with a complementary mechanism: the tendency of people to group together with likeminded individuals to present novel *arguments and information* in favor of their preferred views, leading individuals in the group to become more easily convinced of even more extreme views.⁴² For example, members of different political parties tend to frequent different news sources and social media networks.⁴³ These differential sources of information tend to provide members of each group with new information and arguments supporting their members’ preexisting views, ignoring or delegitimizing countervailing information and arguments, thereby making individuals in each groups more likely to develop more polarized views.⁴⁴ Further, research indicates this mechanism is especially strong for “intellectant” issues—or issues involving intellectual debate, including moral issues.⁴⁵ Informational Influence Theory also coheres with a well-established individual bias: *confirmation bias*, the tendency of people to selectively seek and privilege information confirming their preexisting beliefs, while ignoring or minimizing contrary information.⁴⁶

Finally, Self-Categorization offers a third mechanism for polarization. When individuals are confronted with a risky or threatening outgroup, there is a pronounced human tendency to

⁴⁰ Isenberg (1986).

⁴¹ Yardi and Boyd (2010).

⁴² See Vinokur and Burnstein (1974)

⁴³ See e.g. Iyengar & Hahn (2009).

⁴⁴ See e.g. Morris (2007).

⁴⁵ Isenberg (1986).

⁴⁶ Plous (1993): 233.

coalesce around the views of one's self-identified in-group as a kind of protection-mechanism of solidarity against the threatening out-group.⁴⁷ This general mechanism is familiar from everyday life and history—as when Adolf Hitler used perceived threats (by Jews and others) to rally the German people behind his extreme Nazi ideology, and in polarized debates over moral issues today. For instance, anti-abortionists may be cast proponents of abortion as complicit in “genocide”⁴⁸ whereas proponents of abortion cast anti-abortionists as “waging a war on women.”⁴⁹ Finally, this mechanism appears to strengthen the more threatening the outgroup is perceived to be.⁵⁰

We can now provide several theoretical arguments that Discovery Model likely plays a significant role in group polarization. First, the Discovery Model appears to play directly into the phenomena described by Social Comparison Theory. Once again, Social Comparison Theory shows that people tend to seek approval of those they interact with, adopting progressively more extreme views to impress those in their group. As such, Social Comparison Theory predicts that if people cluster around opposing moral “discoveries”—if, for instance, some believe they have discovered abortion is wrong, whereas others believe they have discovered abortion is permissible—there will be a progressive tendency for each group's members to adopt more extreme positions in order to impress members of their own group. Which, of course, is basically what we do see across a wide variety of moral issues. We see people cluster together in opposing moral groups—pro-abortion and anti-abortion groups, gun-control and gun-rights groups, pro-Israel groups and pro-Palestinian groups, etc.—with members on each side often “ramping up”

⁴⁷ Hogg et al (1990).

⁴⁸ Cunningham (2009).

⁴⁹ Andrews et al. (2017).

⁵⁰ McGarty (1992).

their moral claims to impress fellow group-members.⁵¹ Social Comparison Theory thus not only predicts that the Discovery Model likely plays a significant causal role in polarization. Social Comparison Theory and the Discovery Model together provide a potential explanation for why polarization is so pronounced on “judgmental” issues (specifically, moral issues).⁵² Because moral beliefs involve or are related to *reactive attitudes* such as anger and blame⁵³, when members of opposing moral groups believe they have made opposing “moral discoveries” (viz. the Discovery Model), members of each group may adopt more extreme views in order to appeal to the reactive attitudes of members of their group (playing to their group’s anger, resentment, etc.).⁵⁴ Social Comparison Theory thus predicts not only that the Discovery Model likely plays a causal role in polarization, but that likely plays a prominent role.

Now turn to Informational Influence Theory, the theory which holds that polarization occurs by people in groups providing each other novel arguments and information that *confirm* their pre-existing beliefs, making them more amenable to even more extreme views. Here again, the Discovery Model appears to play directly into these phenomena. Informational Influence Theory predicts that that if people cluster around opposing moral “discoveries”—if, for instance, some believe they have discovered abortion is wrong, whereas others believe they have discovered abortion is permissible—people will tend to provide *new arguments and information* to members of their own favored moral group, making members of each group progressively more amenable to more extreme beliefs. Yet this too is broadly what we see in everyday life. In the abortion debate, for instance, we see many novel philosophical arguments generated on each side of the debate,

⁵¹ Tosi & Warmke (2016).

⁵² See e.g. Haidt (2012) and Isenberg (1986).

⁵³ Strawson (1963).

⁵⁴ See Tosi & Warmke (2016): §2. For examples, see e.g. Cunningham (2009), Rostenberg (2014), and PoliticsUSA (2017).

with each side tending to emphasize the novel arguments for their own moral beliefs while ignoring or dismissively discounting arguments for the other side's beliefs.⁵⁵ This is clearly not an isolated case, as people are known to cluster in political groups around different moral issues—groups which tend to expose their members to different information and arguments.⁵⁶ Further, as we saw earlier, Information Influence Theory predicts that polarization tends to be particularly pronounced for “intellectant” issues.⁵⁷ Because people who subscribe to the Discovery Model commonly treat moral matters as issues of intellectual debate—things to debate at dinner tables, on television, in university seminars, etc.—Informational Influence Theory thus predicts that the Discovery Model likely plays a prominent role in polarization.

Finally, the Discovery Model also appears to play directly into the phenomena described by Self-Categorization Theory. Self-Categorization Theory predicts that group polarization tends to occur when an in-group is confronted by a threatening out-group. It is plain from everyday experience, however, that in-groups and “threatening outgroups” are often defined precisely in *moral terms*—in terms of “moral truths” people believe members of their group to have discovered. For instance, both sides of the abortion debate clearly find the other group threatening⁵⁸—and the same is clearly true across a variety of moral issues, including the Israeli-Palestinian conflict, gun-control, and so on. Insofar as human beings often do identify as members of *moral groups*—people who cluster around similar moral “discoveries”, viewing opposing groups as threatening out-groups—Social Categorization Theory also predicts that the Discovery Model plays a causal role in polarization.

⁵⁵ For instance, whereas anti-abortion websites such as Arthur (2001) and Pro-Life Perspective (2016) tend to emphasize philosophical arguments defending the moral status of human fetuses (viz. Marquis 2007 and Pojman 1998), pro-choice websites such as NLRC (2016) tend to emphasize arguments defending women's rights to their bodies (viz. Thomson 1976).

⁵⁶ See e.g. Iyengar & Hahn (2009).

⁵⁷ Isenberg (1986).

⁵⁸ Again, see Andrews et al. (2017) and Cunningham (2009).

3 An Anti-Polarizing Alternative? The Negotiation Model

I have argued elsewhere⁵⁹ on meta- and normative ethical grounds that although a few moral ideals (of coercion-minimization, mutual assistance, and fair bargaining) can be discovered through rational argumentation, all other moral truths—including how the above ideals should be applied to applied ethical topics—should be thought of not as discoverable through intuition, argument, or any other cognitive or affective process, but instead as *created* by interpersonal negotiation. Let us call this the Negotiation Model of morality.

The basic idea behind the Negotiation Model is straightforward. Consider again the issue of abortion. The Discovery Model holds that there are preexisting facts about the morality of abortion (viz. its rightness, wrongness, etc.) that we can discover through intuition, argument, or some other cognitive or affective process. In contrast, the Negotiation Model holds that the moral status of abortion is literally indeterminate unless and until a social compromise has been arrived at, after which point abortion's moral status should be seen as *defined* by norms negotiated, in essence *settling* abortion's moral status via negotiated compromise (as in: "We have now negotiated a compromise that abortion is permissible in conditions A, B, and C, but impermissible in conditions X, Y, and Z"). On the Negotiation Model, as such, it is a meta- and normative-ethical mistake to form moral beliefs prior to social negotiation. People should instead withhold judgment on controversial moral issues, not forming moral beliefs on those issues until after clear public norms have been negotiated, after which point they should form beliefs in line with the negotiated norms (as in: "I now believe that abortion is morally permissible in conditions A, B, and C, but impermissible in conditions X, Y, and Z, because these are the standards that have been publicly negotiated as a compromise"). Importantly, on the Negotiation Model, these publicly negotiated

⁵⁹ [Reference redacted to preserve anonymized review].

norms—and the moral beliefs they prescribe—are not mere “maxims” or rules to follow for some further moral aim (such as, say, utility-maximization). The norms instead express genuine *moral propositions* about the issue in question (viz. the moral permissibility or impermissibility of abortion) that individuals should *believe*.⁶⁰

The Negotiation Model obviously raises many empirical and philosophical questions. First, is the model psychologically realistic? Can people really believe (for instance) that abortion’s moral status is indeterminate prior to public negotiation, and then come to believe that public negotiation settles its moral status? Second, is the model meta-ethically and normatively justifiable? Can negotiation truly settle the moral status of abortion? Although I have argued elsewhere that the Negotiation Model is indeed justifiable meta-ethically, normatively, and psychologically⁶¹, these are broad issues that we cannot settle here. Instead, let us examine the Negotiation Model’s theoretical relationship to this paper’s topic: group polarization.

On my preferred version of the Negotiation Model, moral truths are created by the outcome of *negotiated agreements* between all agents plausibly motivated by moral ideals of coercion-minimization, mutual assistance, and equal bargaining power.⁶² However, because my favored account is controversial⁶³, let us define the Negotiation Model here more broadly, in terms of the following meta-ethical and psychological claims:

- **The negotiation model of meta-ethics:** aside from perhaps some moral ideals which may be discovered by rational argument (more on this shortly), moral truths do not exist to be discovered by intuition, argument, or any other cognitive or affective process, but are instead *created* by interpersonal processes of moral negotiation that, at the very least,

⁶⁰ I thank an anonymous reviewer for suggesting this clarification.

⁶¹ [Reference redacted for anonymized review].

⁶² [Reference redacted for anonymized review].

⁶³ [Reference redacted for anonymized review].

involve all those with interests in the issue in question *seeking to arrive at a compromise agreement* on moral norms for that issue.

- **The negotiation model of moral-belief formation:** individuals who endorse the negotiation model of meta-ethics should tend to *seek common ground* with others who share relevant moral ideals; believe that morality *requires* moral negotiation as such; and *form moral beliefs* only after negotiating compromises, in line with whatever norms have resulted from compromise agreement.

Because these claims are complex, allow me to spell out their components a bit more.⁶⁴ “Moral negotiation”, as I wish to understand it, is a term of art intended to cover any and all forms of interpersonal human behavior (such as conversation, voting, and so on) that involve those with interests in a given moral issue (abortion, etc.) obeying certain discursive rules—rules that at the very least include a commitment to (A) *certain regulative moral ideals* as background beliefs and motivations, (B) *suspending moral judgment* on the issue in question before interpersonal agreement is reached, and (C) *seeking a compromise agreement* on moral norms for the issue in question (e.g. abortion) with others who plausibly satisfy conditions (A) and (B). Allow me to illustrate using my preferred version of the Negotiation Model—a version which, again, I do not want to presuppose the truth of here, but merely use for illustrative purposes.

As mentioned earlier, I have argued elsewhere that moral ideals of coercion-minimization, mutual assistance, and equal bargaining power *can* be established (i.e. discovered) through rational argument.⁶⁵ However, I argue that these are only regulative ideals—ideals that people can have legitimate interests in weighing against one another. Accordingly, on my favored version of the

⁶⁴ I thank an anonymous reviewer for inviting me to explicate the nature of negotiation and the model’s discursive rules in more detail.

⁶⁵ [Reference redacted for anonymized review].

Negotiation Model, moral truths on controversial moral issues (e.g. abortion) should be understood in terms of the *outcome* of a compromise agreement between all those who plausibly share the above regulative ideals (of coercion-minimization, mutual assistance, etc.), but who may have different priorities on the issue in question. The example of abortion is, I think, instructive here. On my account, a major reason why the morality of abortion remains such a divisive issue is because individuals on both sides of the debate plausibly *share* relevant regulative moral ideals. Anti-abortion advocates, for example, typically claim that fetuses have a “right to life”—a claim clearly intended to *help fetuses*, protecting them from having their lives *coercively* ended. Pro-abortion advocates, on the other hand, claim women have a “right to choose”—a claim clearly intended to *help women*, protecting their reproductive choices from being *coercively reduced*. Both sides are thus plausibly motivated by regulative ideals of coercion-minimization and assisting others. The primary difference between the two sides—on my version of the Negotiation Model, at least—occurs at the level of moral priorities: anti-abortion advocates currently think the rights of the fetus should “trump” (or outweigh) the rights of women, whereas pro-abortion advocates think the rights of women should take priority. On my favored version of the Negotiation Model, because both sides share relevant regulative ideals but have different priorities, they have a duty to negotiate a compromise: it is *wrong* for both sides to think that they can “discover” moral correctness of their own preferred view (e.g. “Abortion is wrong!”) via intuition, argument, or any other process of discovery. Instead, the Negotiation Model requires both sides to both recognize the legitimacy of the other side’s concerns (since both sides plausibly share relevant regulative ideals), and then demonstrate a willingness to forge a compromise agreement.

What kind of negotiated compromise might emerge from such a process, say in the case of abortion? One obvious possibility—but not the only possible one—is this: because prevailing

scientific knowledge indicates fetuses first become sentient between 18 to 25 weeks of gestation⁶⁶, both sides of the abortion debate could (if they were willing to conform to the Negotiation Model) arrive at a compromise agreement that (i) early abortion prior to fetal sentience is morally permissible, (ii) abortion after fetal sentience is normally impermissible, except perhaps in cases of rape, incest, or danger to the mother's life, and finally (iii) members of society share a duty to devote ample social resources to provide sexually-active women with ready access to family-planning resources (including access to affordable early-term abortion) to prevent abortion after fetal sentience.

Such a compromise would almost certainly not fully satisfy many parties to the abortion debate—in part, I think, because the Discovery Model is so deeply entrenched in how people think about moral issues (viz. “But abortion is *murder!*”), but also because of the very nature of compromise (which requires “give and take”). Still, as uncomfortable as compromise may be, if the Negotiation Model is correct (as I have argued elsewhere it is), we should nevertheless accept that it is what morality requires. Further, and importantly, because people might not be wholly satisfied with a given compromise, compromises following the Negotiation Model would plausibly leave many matters open to *renegotiation*. If, for instance, the above compromise on abortion had serious negative effects on women's lives (e.g. by requiring single women to bear children if the father dies late in pregnancy), individuals on the “pro-choice” side of the debate could bring that new information to bear publicly in the aim of renegotiating abortion norms (as in: “I know we have currently agreed that abortion is permissible in cases X, Y, and Z. However, because this standard of permissibility is having negative effects on women, I would like us to consider a new compromise on different standards of permissibility”). In this way, the Negotiation Model

⁶⁶ See Tawia (1992) and Koch (2009).

entails—plausibly, I believe, albeit provocatively—that moral truths on controversial issues can literally *evolve* as people lobby for and negotiate new compromises.

It might seem hard to imagine many people ever accepting the Negotiation Model, as it would require us to give up many moral convictions (about abortion, etc.) that we may believe very deeply. Indeed, the Negotiation Model’s psychological model of moral-belief formation might turn out to be difficult or even impossible for people to reliably conform to, if Haidt’s Social Intuition Model (SIM) of moral-belief formation is correct.⁶⁷ Further, some may worry that the Negotiation Model would have undesirable moral consequences, such as requiring gays and lesbians to negotiate on same-sex marriage—compromises that might set back the gay liberation movement.⁶⁸ All of these are important questions worth investigating in more detail in future research. However, because we cannot settle them here, let us instead examine the Negotiation Model’s theoretical relationship to our primary topic of inquiry: group polarization.

...[Abridged for conference presentation]...

Let us begin with Social Comparison Theory: the theory which holds that group polarization results from people adopting more extreme views to *impress* fellow group members. The Negotiation Model promises to undercut this polarizing mechanism in at least two ways: by (A) discouraging people from adopting first-order moral beliefs (e.g. “Abortion is wrong!”) prior to negotiating, and (B) encouraging people to see answers to controversial moral questions as *created* through negotiated compromise. These two discursive elements of the Negotiation Model promise to defuse polarizing mechanisms of social comparison at “step 1”: it would lead people

⁶⁷ See Haidt (2012). It is important to note here that empirical data supporting the SIM model have been collected under prevailing social conditions—in which most people appear to tacitly or explicitly accept the Discovery Model. Consequently, it is an open question whether a “paradigm shift” in the direction of the Negotiation Model might substantially *change* how individuals form moral beliefs.

⁶⁸ I thank an anonymous reviewer for raising these concerns.

have no settled moral beliefs for groups to cluster around (e.g. “Abortion is wrong”) in the first place. On the contrary, it would plausibly give people an *anti-polarizing* ideal to cluster around despite their differences: the ideal of negotiating compromises with people with different priorities who share relevant moral ideals. Importantly, existing research already indicates that group organization around such a cooperative ideal does indeed mitigate polarization and promote cooperation.⁶⁹ As such, the Negotiation Model theoretically promises to harness the forces that Social Comparison Theory identifies as responsible for group polarization to produce the very opposite: progressive convergence on a cooperative norm of negotiated compromise.

Now consider Informational Influence Theory, the theory which holds that group polarization is generated precisely by people seeking out and attending selectively to information that confirms their preexisting belief. The Negotiation Model promises to undermine the social psychological forces responsible for group polarization here as well. When people have preexisting first-order moral beliefs on a given side of an issue, as on the Discovery Model—such as the preexisting belief that abortion is wrong, or alternatively, that women have a right to abortion—the phenomenon of informational influence leads *each side* to attend to information in support of their preexisting moral beliefs, making each side more amenable to more extreme beliefs. The Negotiation Model, on the other hand, holds that people should *not have* opposing first-order moral beliefs in the first place, but should instead conceive answers to moral questions (such as the morality of abortion) as *created* through negotiated compromise. But now if people increasingly held this kind of first-order moral belief—the belief that answers to controversial moral questions must be negotiated—then Informational Influence Theory predicts that people would become progressively more inclined to seek out and attend selectively to their belief that *moral answers*

⁶⁹ Brewer (1996).

must be created cooperatively through negotiation. Consequently, the Negotiation Model theoretically promises to harness the phenomena described by Informational Influence Theory to *prevent* polarization and promote cooperation.

Finally, consider Social-Categorization Theory, the theory which holds that group polarization results from the development of *in-groups* which then treat out-groups as threatening. As we have seen, in-groups and out-groups often form around divisive moral “discoveries.” For example, whereas anti-abortionists often cast defenders of abortion as a threatening out-group (“They are baby-killers!”), defenders of abortion often cast anti-abortionists the very same way (“They want to take away women’s rights!”). The Negotiation Model once again promises to halt this polarizing force at “step 1.” Insofar as it (A) holds that people should not have settled moral beliefs on issues prior negotiation, and (B) should be willing to negotiate compromises with those with different priorities, the Negotiation Model would theoretically *prevent the formation* of divisive in-groups and out-groups, instead promoting the development of a *cooperative in-group*: people who have different priorities on controversial moral issues (abortion, gun control, etc.), but who are nevertheless unified around anti-polarizing ideals of negotiation and compromise (something which, again, has indeed been found to promote cooperation⁷⁰). Finally, Social-Categorization Theory does plausibly predicts that the Negotiation Model would generate certain types of polarization—namely, polarization between those who accept the Negotiation Model and those who accept the Discovery Model (who might indeed regard each other as threatening “enemies”), as well as polarization with those who reject relevant regulative ideals (e.g. racists, sexists, etc.). However, while Social-Categorization Theory plausibly predicts that there would be polarization between these groups—with each potentially treating the others as threatening out-

⁷⁰ Brewer (1996).

groups—these would not obviously be bad forms of polarization according to the Negotiation Model, as the model itself suggests that we should *not* tolerate the Discovery Model or false regulative ideals.

In sum, all three empirical theories of group polarization predict that the Negotiation Model is likely to substantially reduce polarization relative to the Discovery Model, and perhaps even harness social-psychological forces to generate the opposite: a progressive willingness to cooperate and compromise.

Conclusion

Laypeople and philosophers tend to treat moral truths as discoverable through intuition, argument, or other cognitive or affective process. However, we have seen that there are strong theoretical reasons—based on three empirically-supported theories of group polarization—to believe this Discovery Model of morality is a likely cause of polarization: a social-psychological phenomenon known to have a wide variety of disturbing social effects. We then saw that there are complementary theoretical reasons to believe that an alternative, Negotiation Model of morality might not only mitigate polarization but actually foster its opposite: an increasing willingness for to work together to arrive at compromises on moral controversies. While this paper does not prove the existence of the hypothesized relationships between the Discovery Model, Negotiation Model, and polarization, it demonstrates that there are ample theoretical reasons to believe that such relationships are likely and worthy of further empirical and philosophical research.

References

- Abrams, D.; M. Wetherell; S. Cochrane; M.A. Hogg; J.C. Turner (1990). 'Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization', *British Journal of Social Psychology* 29 (2): 97–119.
- Andrews, B., Levintova, H., Liss-Schultz, N. & P.R. Lockhart (2017). 'Trump Brought the War on Women Mainstream in His First 100 Days', *Mother Jones*, <http://www.motherjones.com/politics/2017/04/first-100-days-reproductive-rights-abortion-trump>, accessed on 6/5/2017.
- Arthur, J. (2001). 'Personhood: Is a Fetus a Human Being', *Pro Choice Action Network*, <http://www.prochoiceactionnetwork-canada.org/articles/fetusperson.shtml>, accessed 3/20/16.
- Aristotle, *Nicomachean Ethics*, T. Irwin (trans.) (1999). *Aristotle: Nicomachean Ethics*, 2nd Ed. (Indianapolis: Hackett).
- Audi, R. (2015). 'Intuition and Its Place in Ethics', *Journal of the American Philosophical Association* 1(1): 57-77.
- Baldassarri, D., & Gelman, A. (2008). 'Partisans without constraint: Political polarization and trends in American public opinion', *AJS: American journal of sociology*, 114(2), 408.

- Barcalow, E. (2007). *Moral Philosophy: Theories and Issues*, 4th ed. (Belmont, CA: Thomson Wadsworth).
- Black, P.; B. Wedeman, & J. Mullen (2015). 'Israeli Palestinian Conflict: Two Viewpoints; One Tragic Outcome', *CNN*, <http://www.cnn.com/2015/10/21/middleeast/israeli-palestinian-root-of-violence>, accessed 6/20/16.
- Bray, R. M. & A. M. Noble (1978). 'Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization', *Journal of Personality and Social Psychology* 36 (12): 1424–1430
- Brewer, M. B. (1996). 'When contact is not enough: Social identity and intergroup cooperation', *International journal of Intercultural relations*, 20(3-4), 291-303.
- Brink, D. (1989). *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).
- Cunningham, G. (2009). 'Is Abortion Genocide?', *Canadian Center for Bio-Ethical Reform*, <https://www.endthekilling.ca/training/classroom/genocide>, accessed 6/2/2017.
- Dancy, J. (2013). 'Moral Particularism', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), <http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>.

----- (1986). 'Two Conceptions of Moral Realism', *Proceedings of the Aristotelian Society*,
Supplementary Volume 60: 167-87.

Daniels, N. (2013). 'Reflective Equilibrium', in E.N. Zalta (ed.), *The Stanford Encyclopedia of
Philosophy* (Winter 2013 Edition),
<http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>, accessed 6/20/16.

Durkheim, E. [1893]. *The Division of Labor in Society*, G. Simpson (trans.) (New York: the Free
Press).

Finlay, S. (2007). 'Four Faces of Moral Realism', *Philosophy Compass 2*: 820-49.

Haidt, J. (2012). *The righteous mind: why good people are divided by politics and religion* (New
York: Vintage Books).

---- (2001). 'The emotional dog and its rational tail: a social intuitionist approach to moral
judgment', *Psychological review*, 108(4), 814.

Hinsz, V.B.; J.H. Davis (1984). 'Persuasive Arguments Theory, Group Polarization, and Choice
Shifts', *Personality and Social Psychology Bulletin*10: 260–268.

Hogg, M.A.; Turner, J.C.; Davidson, B. (1990). 'Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization', *Basic and Applied Social Psychology* 11: 77–100.

Huckfeldt, R., Mendez, J.M., & Osborn, T. (2004). 'Disagreement, ambivalence, and engagement: The political consequences of heterogeneous networks', *Political Psychology*, 25(1), 65-95.

Hursthouse, R. (1999). *On Virtue Ethics* (Oxford: Oxford University Press).

Isenberg, D.J. (1986). 'Group Polarization: A Critical Review and Meta-Analysis', *Journal of Personality and Social Psychology* 50 (6): 1141–1151.

Iyengar, S. & Hahn, K.S. (2009). 'Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use', *Journal of Communication* 59(1): 19-39.

Johnston, L.C. & C.N. Macrae. 'Changing social stereotypes: The case of the information seeker', *European Journal of Social Psychology* 24.5 (1994): 581-592.

Joyce, R. (2015). 'Moral Anti-Realism', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), <http://plato.stanford.edu/archives/sum2015/entries/moral-anti-realism/>.

----- (2007). *The Myth of Morality* (Cambridge, UK: Cambridge University Press).

Kant, I. (1797). *The Metaphysics of Morals*, in M.J. Gregor (ed.), *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy* (Cambridge: Cambridge University Press, 1996): 363-603.

---(1785). *Groundwork of the Metaphysics of Morals*, in M.J. Gregor (ed.), *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy* (Cambridge: Cambridge University Press, 1996): 38-108.

Kaplan, M.F. (1977). 'Discussion Polarization Effects in a Modified Jury Decision Paradigm: Informational Influences', *Sociometry* 3 40(3): 262–271.

Katsafanas, P. (2011). 'Deriving Ethics from Action: A Nietzschean Version of Constitutivism', *Philosophy and Phenomenological Research* 83(3): 620-60.

Koch, C. (2009). 'When Does Consciousness Arise in Human Babies?', *Scientific American*, <https://www.scientificamerican.com/article/when-does-consciousness-arise/>, accessed 6/1/2017.

Korsgaard, C.M. (2009). *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press).

----- (2008). *The Constitution of Agency* (Oxford: Oxford University Press).

----- (2000). *Creating the Kingdom of End* (New York: Cambridge University Press).

Liberty Women's Health (2016), 'Freedom to Choose',
<http://www.libertywomenshealth.com/rtc.php>, accessed 6/20/16.

Luhan, Wolfgang J.; Martin G. Kocher; Matthias Sutter (2009). 'Group Polarization in the Team Dictator Game Reconsidered', *Experimental Economics* 12 (1): 26–41

Mackie, D.M. (1986). 'Social Identification Effects in Group Polarization', *Journal of Personality and Social Psychology* 50: 720–728.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong* (London: Penguin UK).

Marquis, D. (2007). 'An Argument Abortion is Wrong', in Russ Shafer-Landau (ed.), *Ethical Theory: An Anthology* (Blackwell Publishing): 439-50.

McGarty, Craig; John C. Turner, Michael A., Barbara David; et al. (1992). 'Group polarization as conformity to the prototypical group member', *British Journal of Social Psychology* 31: 1–19

Morris, J.S. (2007). 'Slanted Objectivity? Perceived Media Bias, Cable News Exposure, and Political Attitudes', *Social Science Quarterly* 88(3): 707-28.

Myers, D.G.; S.J. Arensen (1972). 'Enhancement of dominant risk tendencies in group discussion', *Psychological Reports* 30: 615–623.

Myers, David G.; G.D. Bishop (1970). 'Discussion effects on racial attitude', *Science* 169 (3947): 778–779

Newman, L.S. (2002). *Understanding Genocide: The Social Psychology of the Holocaust*. New York, NY: Oxford University Press. pp. 143–161

NRLC (2016). 'Mission Statement', www.nrlc.org/about/mission, accessed 6/20/16.

Parfit, D. (2011). *On What Matters*, Vols. 1&2 (Oxford: Oxford University Press).

Pew Research Center (2016). 'A Wider Ideological Gap Between More and Less Educated Adults', <http://www.people-press.org/2016/04/26/a-wider-ideological-gap-between-more-and-less-educated-adults/>, accessed on June 6, 2017.

Plous, Scott (1993). *The Psychology of Judgment and Decision Making* (McGraw Hill).

PoliticsUSA (2017). 'Proof of the GOP War on Women', <http://www.politicususa.com/proof-war-women-2>, accessed 6/5/2017.

Pojman, L.P. (1998). 'Abortion: A Defense of the Personhood Argument', *The Abortion Controversy 25 Years After Roe v. Wade: A Reader* (Cenage).

Pro-Life Perspective (2016). 'About the show', *Pro-Life Perspective*, www.prolifeperspective.com, accessed 6/20/16.

Pruitt, D. (1971). 'Choice shifts in group discussion: An introductory review', *Journal of Personality and Social Psychology* 20 (3): 339–360.

Rapp, C. (2016). 'Moral opinion polarization and the erosion of trust', *Social science research*, 58, 34-45.

Ross, W.D. (1930). *The Right and the Good* (Oxford: Oxford University Press, 2002).

Rostenberg, G. (2014). 'Why Abortion is Genocide', *Live Action News*, <https://www.liveaction.org/news/why-abortion-is-genocide/>, accessed 6/2/2017.

Scanlon, T.M. (1998). *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

Shafer-Landau, R. (2003). *Moral Realism: A Defense* (Oxford: Oxford University Press).

Shoemaker, D.W. (2011). 'Psychopathy, Responsibility, and the Moral/Conventional Distinction', *Southern Journal of Philosophy*, 49(s1): 99-124

- Stoner, J.A. (1961). 'A comparison of individual and group decision involving risk', *Unpublished master's thesis, Massachusetts Institute of Technology*.
- Stratton-Lake, P. (2014). 'Intuitionism in Ethics', in E.N. Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), <http://plato.stanford.edu/archives/win2014/entries/intuitionism-ethics/>, accessed 6/16/2016.
- Strawson, P.F. (1962). 'Freedom and resentment', *Proceedings of the British Academy* 48:1-25.
- Stroud, N.J. (2010). 'Polarization and partisan selective exposure', *Journal of Communication*, 60(3), 556-576.
- Sunstein, C.R. (2002). 'The Law of Group Polarization', *The Journal of Political Philosophy* 10 (2): 175–195
- Tawia, S. (1992). 'When is the capacity for sentience acquired during human fetal development?', *Journal of Maternal-Fetal Medicine*, 1(3), 153-165.
- Thomson, J.J. (1976). 'A Defense of Abortion', in *Biomedical Ethics and the Law* (Springer).
- Timmons, M. (2007). *Disputed Moral Issues: A Reader* (New York: Oxford University Press).

- Tosi, J. & Warmke, B. (2016). 'Moral Grandstanding', *Philosophy and Public Affairs* 44(3): 197-217.
- Van Roojen, M. (2015). 'Moral Cognitivism vs. Non-Cognitivism', in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), <http://plato.stanford.edu/archives/fall2015/entries/moral-cognitivism>, accessed 6/20/16.
- Vaughn, L. (2009). *Bioethics: Principles, Issues, and Cases* (Oxford: Oxford University Press).
- Vinokur, A. & Burnstein, E. (1974). 'Effects of partially shared persuasive arguments on group induced shifts: A group problem-solving approach', *Journal of Personality and Social Psychology* 29: 305–315.
- Walker, T.G. & Main, E.C. (1973). 'Choice shifts and extreme behavior: Judicial review in the federal courts', *The Journal of Social Psychology*. 2 91 (2): 215–221.
- Warren, M.A. (1973). 'On the Moral and Legal Status of Abortion', *The Monist*: 43-61.
- Yardi, S. & Boyd D. (2010). 'Dynamic Debates: An analysis of group polarization over time on Twitter', *Bulletin of Science, Technology and Society* 30 (5): 316–27.