Moral alchemy: How love changes norms

Rachel W. Magid & Laura E. Schulz

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Cambridge, MA 02139 USA

Address correspondence to:

Rachel Magid

rwmagid@mit.edu

Department of Brain and Cognitive Sciences

Massachusetts Institute of Technology

77 Massachusetts Ave, 46-4011

Cambridge, MA 02139

Abstract

We discuss a process by which non-moral concerns (that is concerns agreed to be non-moral within a particular cultural context) can take on moral content. We refer to this phenomenon as *moral alchemy* and suggest that it arises because moral obligations of care entail recursively valuing loved ones' values, thus allowing propositions with no moral weight in themselves to become morally charged. Within this framework, we predict that when people believe a loved one cares about a behavior more than they do themselves, the moral imperative to care about the loved one's interests will raise the value of that behavior, such that people will be more likely to infer that third parties will see the behavior as wrong (Experiment 1) and the behavior itself as more morally important (Experiment 2) than when the same behaviors are considered outside the context of a caring relationship. The current study confirmed these predictions.

Moral alchemy: How love changes norms

Moral dilemmas in psychology play a critical role in probing our intuitions and revealing the complexities underlying our moral judgments. In the interest of understanding the foundations of moral reasoning, people have been asked if it is okay to sacrifice one person to save five (e.g., Cikara, Farnsworth, Harris, & Fiske, 2010; Crockett, Clark, Hauser, & Robbins, 2010; Cushman, 2015; Foot, 1967; Mikhail, 2000; 2007), accept stolen goods (Haidt, 2007; Haidt & Graham, 2007) burn, poison, or shock someone (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolon, 2014; Cushman, 2008; Young, Cushman, Hauser, & Saxe, 2007), have sex with siblings or dead chickens (Graham, Haidt. & Nosek, 2009; Haidt, 2001; Haidt, Bjorklund & Murphy, 2000; Haidt, Koller & Dias, 1993; Prinz, 2006; Young & Saxe, 2011), smother babies (e.g., Greene, Nystrom, Engell, Darley, & Cohen, 2004), eat dead pets (Wheatley & Haidt, 2005), steal drugs (Kohlberg, 1969), harm the environment (Knobe, 2003; 2004; Malle, 20004), smash plates (Piaget, 1932), yank hair (Blair, Marsh, Finger, Blair & Luo, 2006; Nichols, 2002; Nucci, 2001; Turiel, 1983), push someone downhill (Hamlin, Wynn, & Bloom, 2007; Kuhlmeier, Wynn & Bloom, 2003), and desecrate the flag (Gray & Ward, 2011; Haidt et al., 1993). Scenarios like these have revealed surprising subtleties and dissociations in our moral reasoning. Thanks to such thought experiments, we know that the purview of moral reasoning includes not just considerations of harm and fairness, but considerations of authority, loyalty, and purity (Haidt, 2001; see also Blair et al., 2006; Blair, 2009; Haidt & Graham, 2007; Haidt & Joseph, 2007; Rai & Fiske, 2011; Shweder & Haidt, 1993; Shweder, Much, Mahapatra, & Park, 1997). We know that judgments of intentionality differ for violations of harm and violations of purity (Young & Saxe, 2011), differentially influence our intuitions about blame and punishment (Cushman, 2008), and change depending on the causal structure of morally significant events

(Knobe, 2003; 2004; Mikhail, 2000; 2007). Moral thought experiments have furthered our understanding of the early development (e.g., Hamlin & Wynn, 2011; Hamlin et al., 2007; 2010) and neural bases (Crokett et al., 2010; Decety, Michalska, & Kinzler,, 2012; Greene, et al., 2004; Young, et al., 2007; Young, et al., 2010; Young & Dungan, 2012; Young & Saxe, 2008; 2011) of moral reasoning, and have launched vigorous debates on the relative contributions of judgments believed to be rapid, automatic, and affective and those believed to be slow, effortful, and cognitive (see Cushman, 2015 for review).

The extent of these contributions to the psychology of moral reasoning is perhaps the more striking because the moral scenarios that enabled them are, prima facie, remote from human psychology. Most of us will live all our lives without encountering anything very like the dilemmas above. We do of course enact decisions which trade off the good of a few against the good of many, engage in sexual behaviors others might deem perverse, subordinate the needs of infants to other goals, exploit animals and the environment, engage in economic injustices, commit acts of physical aggression, and behave irreverently and disrespectfully. However, to the degree that we worry about such things, we are generally not worried about what to do but about the fact that we do what we shouldn't; the most disturbing aspect of real-world analogues of these scenarios may be our capacity for indifference (Singer, 1972). As scientists, the thought experiments are satisfying because they reveal the paradoxes and ambiguities lurking beneath our moral certitudes. Arguably however, these scenarios reveal the precarious foundation of our moral convictions while leaving our moral anxieties untouched. This is not to say that people do not also sometimes confront ethical challenges with imagination and courage (a topic of psychological and philosophical inquiry in its own right; Anderson, 1999; Railton, 1986; Singer,

1981) but this too arguably contrasts with the moral quandaries that preoccupy us the rest of the time.

Moral alchemy

Here we are interested in "the rest of the time": times when we experience neither moral conviction nor moral complacency, although the stakes (in comparison to the scenarios above) are relatively low. We suggest that the scenarios we *experience* as moral dilemmas do not typically involve questions of intentionality, or pressing conflicts between utilitarian and deontological ends. Rather we believe that many of our everyday moral anxieties center on cases where there is a conflict between our belief in any proposition (including morally neutral ones) and our belief that actions consistent with that proposition will upset someone we love. It is in this sense that love can lead to what we will call *moral alchemy*: caring for others (and indeed the moral obligation to do so) allows propositions with little or no moral weight in themselves to become morally charged. To be very clear, our hypothesis is distinct from the claim that our moral values depend on the values of our close others; many researchers have investigated the degree to which our sense of moral value is affected by moral contagion, or social affiliation (see e.g., Eskine, 2013; Haan, Smith, & Block, 1968; Haidt & Hersh, 2001; Hoffman, Wisneski, Brandt, & Skitka, 2014). Here we are interested in cases where although our own opinion about the actual rightness or wrongness of the behavior may remain unchanged, we nonetheless assign the behavior an elevated moral status.

We will start with a trivial example: the moral status of Pogs. (For those of you who were neither a parent nor child in the 1990's, Pogs are collectible colored disks, originally from bottle caps.) Clearly in the world at large, if someone steps on a Pog, uses one to prop up a table leg, or publically disparages them on national TV, he is morally blameless. He is morally blameless

even if he knows that Pogs are valued by millions of school children in his culture. Suppose however, your child comes up to you and says, "Pogs are the best thing ever." Most of us would be (morally) appalled if you replied, "Pogs are stupid" and snapped a Pog in two.

Of course what is bad in this example is hurting your child's feelings, not hurting Pogs. Nonetheless, we suggest that the effect of moral alchemy is to (locally) change the moral status of Pogs. You cannot disregard them as objects worthy of care and attention without insufficiently valuing your child's values. Critically however, and in contrast to other arbitrary objects that attain moral significance through their association with culturally important moral values (Moll & Schulkin, 2009; Shweder et al., 1987; Turiel, Killen, & Helwig, 1987), Pogs are not valuable because of a symbolic connection to other core values; nor did you reclassify Pogs as agents (as fetuses and non-human animals may be classified; Brandt & Rozin, 2013; Singer, 1995). Pogs did have social-conventional value (for children in the '90's, as a kind of currency) but that is irrelevant here; assuming Pogs have no social currency in 2016, snapping a Pog in front of your Pog-smitten child is still egregious. All that matters is that you knew he cared about Pogs and you did not take his utilities as your own. Note that this is neither moral contagion nor moral duplicity: you do not adopt your child's attitude of valuing Pogs for their own sake but neither do you merely act "as if" you care about Pogs when you do not. Rather, insofar as, and for as long as, failing to care about Pogs would be hurtful to your child, you represent Pogs as objects worthy of care (e.g., you would likely feel guilty about intentionally destroying a Pog, even in private).

Of course many morally neutral things can take on moral content in specific contexts. Basement stairs for the parents of toddlers, or earthquakes for residents of the Pacific Northwest, can be morally relevant insofar as failures to attend to them appropriately could cause harm (and

subsequent guilt). Critically however, stairs and earthquakes don't lose (and may even increase) their moral relevance if the potential victims are indifferent or oblivious to the risk: stairs are intrinsically dangerous to toddlers and earthquakes to Oregonians. Although care for others can make many things, innocuous in themselves, an appropriate target of our moral anxieties, here we reserve the term moral alchemy for transformations of non-moral to moral content that depend solely on others' mental states. Because such transformations require insight into others' unique goals, preferences, values, and beliefs, and because only mental state dependent harms are possible candidates for moral transformation, we believe these are particularly important with respect to moral learning.

Why important? It is after all, uncontroversial that people value idiosyncratic things and that morality requires respecting things that others value. However, we suggest that taken together, these commonplaces of human psychology play a key and under-appreciated role in real life moral dilemmas, moral learning and moral change. Consider a proposition less trivial than "Pogs are the best thing ever." Consider "Academic achievement is important." For the sake of argument, let's presume that within a given cultural context, this counts as a value but not a moral one: everyone concerned accepts that mediocre students can be morally unimpeachable. Suppose however, that your parents are among those who care about this (non-moral) value. If you under-achieve in school, rip up your homework, and refuse to study for tests, are those moral transgressions or not?

We would contend that although the proposition "Academic achievement is important" has no moral content, the proposition "My parents value academic achievement" does. Insofar as your parents may find your actions hurtful and disrespectful to them because you did not take their utilities as your own, a moral issue is at stake. The effect is (loosely) analogous to the

referential opacity induced by complement structures in language: much as the truth value of "It is raining" is independent of the truth value of "Sally believes 'It's raining'", knowing that "My parents care about academic achievement" may have a moral status independent of the moral status of the academic achievement they care about.

We have stressed the importance of close interpersonal relations. Why should it matter that these interactions occur in the context of loving relationships? Why morally, should it matter, that your child cares about Pogs, or your parents care about academic achievement, if, in the world in general, these are largely matters of indifference? We suggest that this is because moral alchemy is only possible when there is a risk of hurt, harm, and interpersonal conflict. If a proposition has moral content in itself (e.g., the belief that "homosexuality is wrong") then moral values (fairness, loyalty, autonomy, care, liberty, purity, etc.) apply broadly; if our parents believe homosexuality is wrong, and we are gay, we may be in trouble simply because one set of moral values (e.g., autonomy, liberty) conflicts with another (care, authority, loyalty). However, propositions without moral content ("Pogs are great", "Academic achievement is important") can take on moral content only in the context of possible violations of care; for matters of indifference to the world at large, this can only happen in the context of loving, intimate relationships.

All of the above requires substantial unpacking, in particular to note the ways in which this idea is distinct from a number of other ideas to which it is, nonetheless, indebted. First, notwithstanding our emphasis on concern for others' feelings, our topic is orthogonal to debates about the relative contribution of emotion (Blair, 1995; Damasio, 1999; Decety et al., 2012; Greene & Haidt, 2002; Shweder & Haidt, 1993; Valdesolo & DeSteno, 2006; Wheatley & Haidt, 2005) and cognition (Kohlberg, 1969; Lombrozo, 2009; Mikhail, 2000; Piaget, 1932; Turiel,

1983) to moral judgment; by the same token, it is orthogonal to syntheses of these views through dual-systems approaches (Campbell & Kumar, 2013; Cushman, Young & Hauser, 2006; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Greene et al. 2004). We are interested in cognition and emotion but rather than considering how first-person emotional reactions to events (e.g., feelings of disgust or shock) either inform or are informed by our cognitive appraisals, we are interested in how valuing others' emotional reactions to events changes the moral status of those events.

Second, in emphasizing caring about others, we are not suggesting that moral cognition reduces to concerns about care and harm. Rather, we believe that obligations of care can give morality a reach that extends beyond the scope even of pluralist taxonomies of morality. That is, through care of others, we can be morally preoccupied by issues that are not intrinsically harmful, and that also may have no bearing on autonomy, community, and divinity (Shweder et al., 1997), reciprocity, purity, hierarchy, and loyalty (Haidt & Graham, 2007); disgust, social conventions, and reciprocity (Blair et al., 2006; Blair, 2009); or unity, hierarchy, equality, and proportionality (Rai & Fiske, 2011).

Finally of course, there is nothing new in the proposal that moral cognition is connected to attachment, kinship and empathetic concern for members of one's own social groups (e.g., Brewer, 1979; Barragan & Dweck, 2014; Rai & Fiske, 2011; Goodall, 1986; Kinzler, Dupoux, & Spelke, 2007; Moll & Schulkin, 2009; de Waal, 1982; 1996; de Waal & Lanting, 1997). In this vein, researchers have suggested that mechanisms evolved for attachment to particular others have allowed people to extend attachment to abstract values, reducing interpersonal conflict and promoting social cohesion (Moll & Schulkin, 2009). Some researchers have gone further, contesting the notion that there is any abstract, universal content to moral judgments at all and

proposing that all moral judgments depend on people's social roles and the interpersonal consequences of those judgments for regulating social relationships (Rai & Fiske, 2011). Other researchers have argued that relationist ethics (or "ethics of care") have been subordinated to analytical ones primarily because of gender biases (Gilligan, 1982; though see also Harding, 1987 on the suspicious tendency to attribute relationist ethics to disempowered social groups).

We are sympathetic to the concern that relationist motives are in tension with universal moral values (e.g., Kant & Gregor, 1988; Rawls, 1971) but not to the tension between "ethics of care" and analytic rational thought. We suggest that the ability to care about relationships is predicated on, not in opposition to, abstract "analytical" cognition. Specifically, we assume that moral reasoning is supported by our ability to reason about others' beliefs and desires, represent others' utilities, and recursively link our utilities to theirs. To the extent that our utilities depend on advancing another's, we will act in their best interests, including promoting the values they care about most (see Kleiman-Weiner & Tenenbaum this issue; see also Jara-Ettinger, Tenenbaum & Schulz; 2015; Jara-Ettinger, Gweon, Tenenbaum & Schulz; 2015; 2016; Ullman, Baker, & Macindoe, Evans, Goodman, & Tenenbaum, 2010; Yoshida, Dolan, & Friston, 2008). Indeed, it is precisely the recursive act of taking another person's utilities as one's own that allows the moral obligations entailed by caring to expand the space of potential moral concerns.

Consequences of moral alchemy

What are the implications of moral alchemy for real world moral dilemmas? When we think of moral failings we are prone to consider problems caused by doing immoral things (stealing, lying, cheating, etc.) or failing to do moral ones (helping the afflicted, preventing harm, etc.). As above, we may also think of the difficulties posed when one moral value conflicts with another (e.g., caring versus fairness or autonomy versus loyalty). These concerns are real

enough. However, we suspect that everyday moral difficulties are posed as often by our desire to do, or not do, something of no particular consequence in itself that takes on moral ramifications only because someone close to us cares about it. When we worry about the heavy-handed edits we made on a colleague's paper, having to miss our daughter's pancake breakfast, or forgetting that our spouse asked us to stop for groceries, we do not merely worry that our choices may cause interpersonal conflict; we worry that we are behaving badly. We worry about this despite fully recognizing that edits, pancake breakfasts, and groceries are morally inconsequential in themselves. They matter only because, and to the degree that, someone close to us cares about them. Absent that caring, anything we did with respect to these would be permissible, and nothing would be forbidden or obligatory. Thus note that moral alchemy does not predict a change in our sense of core moral values: we do *not* come to believe that a given behavior is more wrong because our loved one believes it is more wrong. Rather, moral alchemy predicts that we come to believe that we are more at risk of harming others by treating the behavior as morally neutral; that is, we become more likely to accept that third parties may regard this behavior as morally important.

In the respect that the rightness or wrongness of these actions depends on whether someone cares about them, alchemical norms have something in common with social-conventional norms: we may believe it is right to hang up our backpacks and wrong to leave them in our lockers if this is an expectation but the expectation can change if everyone (or an acknowledged authority) agrees to change it (e.g., Nucci & Turiel, 1978; though see Kelly, Stich, Haley, Eng, & Fessler, 2007 and Nichols & Fold-Bennett, 2003 for discussion). However, in contrast to social-conventional wrongs, in moral alchemy a genuine moral issue is at stake: failing to care about what a loved one cares about can hurt them, at least insofar as hurt feelings

count as “hurt”. Moreover, what is at stake is exactly not conventional. It is personal and idiosyncratic: we can incur obligations to anything the people we love care about, for as long as they care. As a consequence, the world can become something of a moral mine field. We may often feel morally obliged to take actions (or refrain from acting) against our preferences, on issues we personally believe have no moral consequence in their own right, simply because we feel that doing otherwise would hurt someone else.

Thus perhaps it is unsurprising that everyday moral problems frequently take the form of considering whether other people have a right to care as much as they do. Suppose for instance, my own reward in having a clean house, together with my recursive value in promoting my partner’s goals, fail to overcome the costs of cleaning the house. When my partner comes home to a mess, she may reasonably infer that I did not put much weight on her happiness. However, I may believe that she is wrong to value cleanliness so much (or my costs so little). When we experience moral anxiety we often vacillate between guilt that we did not promote another’s utilities and resentment that they have the utilities they do. Consistent with this even young children (and children with autism) show less empathy for others’ distress when they have reason to feel the distress was unreasonable (Leslie, Mallon & DiCorcia, 2006). Note that this tension does not arise in cases where one’s own genuine sense of right and wrong changes because we have changed our minds in response to our loved one’s values. It arises only in the moral alchemy case, where we simultaneously retain our independent estimate of the moral value of a behavior (i.e., as negligible) and a sense that the behavior may nonetheless matter deeply to third parties.

We suggest that the phenomenon of moral alchemy matters to human psychology not only because it may be the source of much of our moral anxiety but also because it is a potential

route to moral change. Consider again the belief that "homosexuality is wrong." As noted, this belief has moral content all by itself; it does not become moral merely because a close other believes it. However, although moral alchemy cannot make an already moral matter moral, it can alter the stakes. One can feel righteous about standing up to homophobia and still feel guilty for grieving and upsetting one's parents. However, if you are gay and your parents believe homosexuality is wrong, you may have an option other than being torn between values of autonomy and liberty and values of caring, loyalty and authority: You can make the case that your parents are wrong to endorse the belief as much as they do. If your parents cease to care intensely about this belief, you can be a gay daughter, and a loyal, loving one. Of course the actual historical context in which beliefs around homosexuality change is far more complex. In the past decades it has involved, among other things, concerns of harm invoked by an epidemic, arguments about intentional choice, issues of equity, and committed activism. However, when many individual people stopped caring as much about homosexuality, many people who had long since concluded that their behavior was, in itself, morally permissible, were freed from the very real moral concern that they nonetheless threatened the happiness of their families. Similar issues are being negotiated today with respect to attitudes towards gender identity.

In addition to playing a role in real world moral dilemmas and moral transformation, we believe that moral alchemy has implications for moral learning. Developmental psychologists have long looked at the effects of parenting styles and particular attitudes and beliefs expressed by parents on children's moral behavior (Baumrind, 1986; Eisenberg & Valiente, 2002; Hoffman, 1970, 1975; Grusec, Goodnow, & Kuczynski, 2000; Kochanska, 1997; Stayer & Roberts, 2004), proposing that children experience "parents as moral philosophers" (Brody & Shaffer 1982, p. 58) and that parental explanations of why actions are right or wrong facilitate

moral development (Kuczynski, 1982; Zahn-Waxler & Chapman, 1982; though see Harris, 1995; 2011). Additionally, studies suggest that children's moral reasoning can be modified by exposing them to adult models whose reasoning differs from their own (Bandura & McDonald, 1963; Brody & Henderson, 1977; Cowan, Longer, Heavenrich, & Nathanson, 1969; Dorr & Fey, 1974).

In light of this emphasis on moral learning, and considerable evidence that pro-social and empathetic concerns are innate or very early-emerging (Aknin, Hamlin, & Dunn, 2012; Hamlin, 2013; Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007; 2010; Joyce, 2006; Katz, 2000; Martin & Clark, 1982; Mikhail, 2011; Sagi & Hoffman, 1976; Warneken & Tomasello, 2007) it is perhaps surprising that children as old as six have a holistic, non-specific view of goodness, clustering positive traits including being hardworking, attractive, intelligent, athletic, kind, artistic and helpful, together. After observing evidence, for instance, suggesting that someone is smart, children often assume they are also helpful; likewise they conclude that someone criticized for a poor artwork is "not nice" (Heyman, Dweck, & Cain, 1992; see also Benenson & Dweck, 1986, Cain, Heyman, & Walker, 2006; Heller & Berndt, 1981, Mullener & Laird, 1971; Stipek & Mac Iver, 1989).

Arguably, children may simply subscribe to a broader morality than adults do. Evidence that children distinguish arbitrary conventions from morals (e.g., Nucci & Turiel, 1978) does not speak to how broadly children may construe moral values themselves. Like the ancient Greeks, they may take all traits that constitute a "good" life as evidence of virtue (Aristotle., & Sachs, 2002). Alternatively, children may ascribe to broad definitions of goodness because adults deliberately foster behaviors (e.g., diligence, orderliness, and self-control) that make individuals better people with whom to have relationships (e.g., Maccoby, 1992); such virtues may become

part of what children believe it means to be a good person, even though adults may not construe these behaviors as moral in the way that avoiding directly harming someone would be.

However, we suggest that children may subscribe to a broad notion of "goodness" because they correctly ascertain that the people they care about most deeply, care deeply about many things, including being hardworking, attractive, intelligent, athletic, kind, artistic and helpful. Although some of these are moral values simpliciter and some are not, insofar as they are valued by the people children love, they may all become subject to moral alchemy. On this account, it is not just that parents use the words "good" and "bad" polysemously, to refer to both moral and non-moral matters, but that non-moral matters become moral ones when they matter to the people you care about. In such contexts, children may come to perceive these as values widely shared by third parties. The situation is made the more complex because the parents' utilities are also recursive: parents often value what they do because they believe it is in the best interest of their children. However, children may fail to understand this (or understand but disagree that their parents' wishes are in their own best interests) while still recognizing that failure to respect and promote their parents' utilities will upset the parent and that this is, in itself, a moral failing.

Moral alchemy experiments

Of course the heart of our contention is that this kind of slippage between moral and non-moral concerns is not limited to moral reasoning in early childhood, it is a feature of every close relationship; thus indications that we moralize non-moral concerns when others care deeply about them should be manifest even in adulthood. The current study is a preliminary test of this claim. In Phase 1 of Experiment 1, we asked two groups of participants to rate how much they cared about items in 21 different categories ranging from matters of dress to matters of harm and

welfare (adapted from previous work on moral reasoning). This measure is intended to get a sense of participants' actual moral judgments: those values they do and do not hold. In Phase 2, we asked them to rate how much either someone they loved (Close Other condition) or an acquaintance (Distant Other condition) would care about a different set of items from each category. In Phase 3, participants were asked to judge the wrongness (on average, for people in general) of failing to do other behaviors drawn from each category. We predicted that for categories where participants judged the other person as caring more than they did (i.e., where the Phase 2 rating was higher than the Phase 1 rating), participants in the Close Other condition would judge that third parties would see failing to engage in those behaviors as more wrong than participants in the Distant Other condition. As discussed, we focus on views about third party judgments because the moral alchemy hypothesis predicts that having close others care more about a value than oneself changes the extent to which we see the value as having importance to third parties, without changing the person's own estimate of the behavior. The moral alchemy hypothesis suggests that this shift is due to the perception that failing to prioritize the values of close others could do harm, rather than to merely being aware that third parties may have different values than oneself. Thus, critically, we both predict that the effect should be specific to the Close Other condition and also that this effect should be specific to cases where participants believe a close other cares *more* (rather than less) than themselves, thus there should be no difference in wrongness judgments between conditions for the remaining items (i.e., categories where the Phase 2 ratings were lower than or identical to the Phase 1 ratings).

In Experiment 2, we replicated Experiment 1, except that instead of asking participants to consider how wrong a behavior was, we asked them to place behaviors on a sliding scale were norms were the least important, values moderately important, and morals most important. We

predicted, that relative to participants in the Distant Other condition, participants in the Close Other condition would "shift right" on this scale, elevating the importance of just those categories of behavior about which the other cared more.

We note, pre-emptively, that insofar as the feelings of close others influence moral judgment, one might expect that effect to obtain broadly, regardless of whether one is asked to think about a close other or not. Additionally we recognize that relative shifts in people's permissibility or value judgments are not tantamount to transforming non-moral values into moral ones. Nonetheless, in the context of a survey-based laboratory task, using items potentially relevant to all participants but not tailor-made to any individual, we believe it would be compelling if considering another person's investment in different kinds of behaviors was associated with a shift in people's value judgments when, and only when, they both cared about the other person, and believed the other person cared more about those behaviors than they did.

Finally, because it is a relatively subtle distinction, we stress that the moral alchemy hypothesis is distinct from the idea that people might share values with our close others, either because people choose to affiliate with those who share their values (e.g., Buss, 1985) or because close others influence each other (e.g., Whitbeck & Gecas, 1988). Rather, the moral alchemy hypothesis predicts that when a loved one values a behavior, it can raise an individual's estimate of the moral status of the behavior, independent of whether it makes the individual herself care about the behavior more. That is, if my loved one cares passionately about academic achievement, homosexuality, veganism, or anything else, I may both A) retain my own independent assessment that these are not morally important issues (i.e., I may believe my loved ones are wrong) and B) nonetheless, elevate my estimate of how third parties will regard the moral status of these topics. Our two different measures ("How strongly would people in general

endorse the statement (that X is wrong)" and "Think about how most people would feel (about whether X is a convention, value, or moral)" are specifically designed to allow for the possibility that people might make one moral judgment for themselves (Phase 1) and a different moral judgment when asked to view the behavior from a third party perspective (Phase 3).

Experiment 1

Methods

Participants.

Participants were recruited on Amazon Mechanical Turk and paid for their participation. One group of participants ($N = 46$, $m_{age} = 33.96$ years, 52% female) was used to norm the stimulus items. Another group of participants ($N = 298$) was recruited for the experimental conditions. Participants were excluded for 1) having previously participated in the norming study or another HIT posted for the study ($n = 32$); 2) failure to answer attention check and comprehension questions correctly ($n = 47$), or 3) a mismatch between their initial and final response in identifying the target person ($n = 5$) (see Inclusion Criteria below for details), resulting in a final sample of 214 participants ($n = 108$ in Close Other condition, $m_{age} = 34.04$ years, 55% female; $n = 106$ in Distant Other condition, $m_{age} = 35.57$ years, 50% female).

Materials.

Forty-eight items in this study, three from each of 16 categories, were adapted from the Moral Foundations Questionnaire (Graham et al., 2009), the Schwartz' Value Scale (Schwartz, 1992), the Portrait Values Questionnaire (Schwartz, Melech, Lehmann, Burgess, Harris, & Owens, 2001), and the European Social Survey (Davidov, Schmidt, & Schwartz, 2008). An additional 15 items from 5 categories (Art and Aesthetics, Dress, Organization and Neatness, Religion, Scholarship) were added by the authors to ensure a diverse range of content. In Phase 1

and 2 of the experiment, items were phrased as statements endorsing beliefs and actions (e.g., "*People should be orderly in their personal space, keeping their homes and offices neat.";* *"People should get as much education as possible."*) In the norming study, and in Phase 3, statements were rephrased if necessary so that participants' endorsements reflected how wrong it would be for people not do something (e.g., "*It is wrong for people not to be orderly in their personal space by keeping their homes and offices neat."*; "*It is wrong for people not to get as much education as possible.")*. See Appendix for a full list of the 63 items in each of the 21 categories.

In norming study, half of the participants rated all three items from ten categories while the other half rated all three items from the other eleven categories. Items were presented in a fixed random order, and participants were also asked to rate how easy the question was to read. In the experimental conditions, a single item was taken from each category to generate three different stimulus sets: A, B, and C. The presentation of items within a set was randomized. A sixth of participants were assigned to set A, B, C, a sixth to A, C, B, a sixth to B, A, C, etc., in Phases, 1-3 respectively. Thus each item occurred in each phase an equal number of times and no individual saw the same item twice.

Procedure.

Stimulus norming.

Participants were tested online using the Qualtrics survey program. They were shown a sliding scale (see example in Figure 1) that they could manipulate with their computer mouse. They were told that they would be given a series of statements and asked to judge how much they agreed with each statement "*Along a sliding scale of endorsement, from 0 (Not at all) to 100 (Passionately). You can place the pointer at any point along the scale."* They were also told,

"*Although these statements are about actions and beliefs, we are not asking how much you personally agree with each statement. Rather think about how most people would respond and use the pointer to rank the statements accordingly. So when responding using the sliding scale, you do not need to consider how you personally feel about the statement in order to answer."* Participants were shown five pictures of the scale with, each with the pointer at a different location, to indicate they could place the pointer at any point on the scale.

After participants rated their endorsement of each statement, they were asked to rate how easy each statement was to read on a Likert scale where 1 indicated *Very Easy* and 7 indicated *Very Difficult*. Only a single statement appeared on the screen at a time. After the participant provided both the endorsement and readability rating, she clicked "Continue" and a new screen appeared with the next statement.

The norming study was used to establish 1) that participants understood the task and used the sliding scale appropriately with respect to common sense judgments of more and less severe wrongs; 2) that there was variability both within and across items in participants' responses, and 3) that the items in each category were easy to read. All of these results were confirmed. Participants used the scale in a meaningful way. For instance an intuitively minor ("*It is wrong not to wear clothes that are appropriate for the season"*), a moderately bad (e.g., *"It is wrong not to show what one is capable of …"*) and a very bad ("*It is wrong to kill a human being"*) offense received mean ratings of 32.80, 56.98, and 77.56 respectively. The categories are listed by rank order from most to least wrong in Table 1, Column 1. There was variability across items (overall, mean: 53.76 (*15.34*); range 22.48-88.90) and the average standard deviation for individual participants' responses within items was 22.61 points, suggesting that there was sufficient variability in participants' perception of how wrong most people would consider any

given norm to test our hypothesis. Overall, the readability of the items were rated an average of 2.36 (.71), or between *Easy* and *Somewhat Easy.*

Experimental design.

Phase 1.

Participants were also tested online using the Qualtrics survey program. They were introduced to the sliding scale as in the norming study. In Phase 1, the items were phrased to reflect endorsement of positive statements (see Appendix, part 1) rather than judgments of wrongness. Participants were instructed as follows: "*For the first part of the study, you will be presented with a number of statements. For each statement, we will ask you to respond on a scale of 0 (not at all) to 100 (passionately) "How strongly do you endorse each statement? Feel free to use the full scale to represent differences in how much you believe in each statement.*" As in the norming study, only a single statement appeared on the screen at a time. After the participant provided a rating, she clicked "Continue" and a new screen appeared with the next statement.

Phase 2.

Phase 2 was similar to Phase 1 except participants were given new items from each category (see Appendix) and asked to rate the items from the perspective of another person. The instructions provided before Phase 2 differed by condition as follows:

Close Other Condition: *For the next part of the study, please first think about another adult who you love and care deeply about. Who are they (e.g., friend, parent, grown child, grandparent, lover, partner, etc.)?* [Enter relationship in a text box.] *Now for the next set of statements, think about the person you love and tell us on a scale of 0 (not at all) to 100*

(passionately) how strongly you think that person would endorse each of the following statements.

Distant Other Condition: *For the next part of the study, please first think about another adult who is a casual acquaintance of yours, someone you know slightly but have no special relationship with. Who are they? (e.g., local merchant, neighbor, distant coworker, etc.)?* [Enter relationship in a text box.] *Now for the next set of statements, think about the person you know and tell us on a scale of 0 (not at all) to 100 (passionately) how strongly you think that person would endorse each of the following statements.*

On each trial in Phase 2, participants in the Close Other condition were asked, "How strongly would the person you love endorse this statement?" and participants in the Distant Other condition were asked, "How strongly would the person you know endorse this statement?" Note that our interest was not in how accurate people were in their ratings of others, but in whether participants' beliefs about the endorsements of close and distant others affected participants' permissibility judgments.

Phase 3.

Phase 3 used a third stimulus set (see Appendix) and was identical to the norming study except that participants were not asked to rate the readability of each statement. Because participants were already familiar with the sliding scale, the instructions were simply: "*You will be asked to respond to the questions along a sliding scale of endorsement, from 0 (Not at all) to 100 (Passionately). You can place the pointer at any point along the scale. Although these statements are about actions and beliefs, we are not asking how much you personally agree with each statement. Rather think about how most people would respond and use the slider to rank*

the statements accordingly. So when responding using sliding scale, you do not need to consider how you personally feel about the statement in order to answer." See Figure 1.

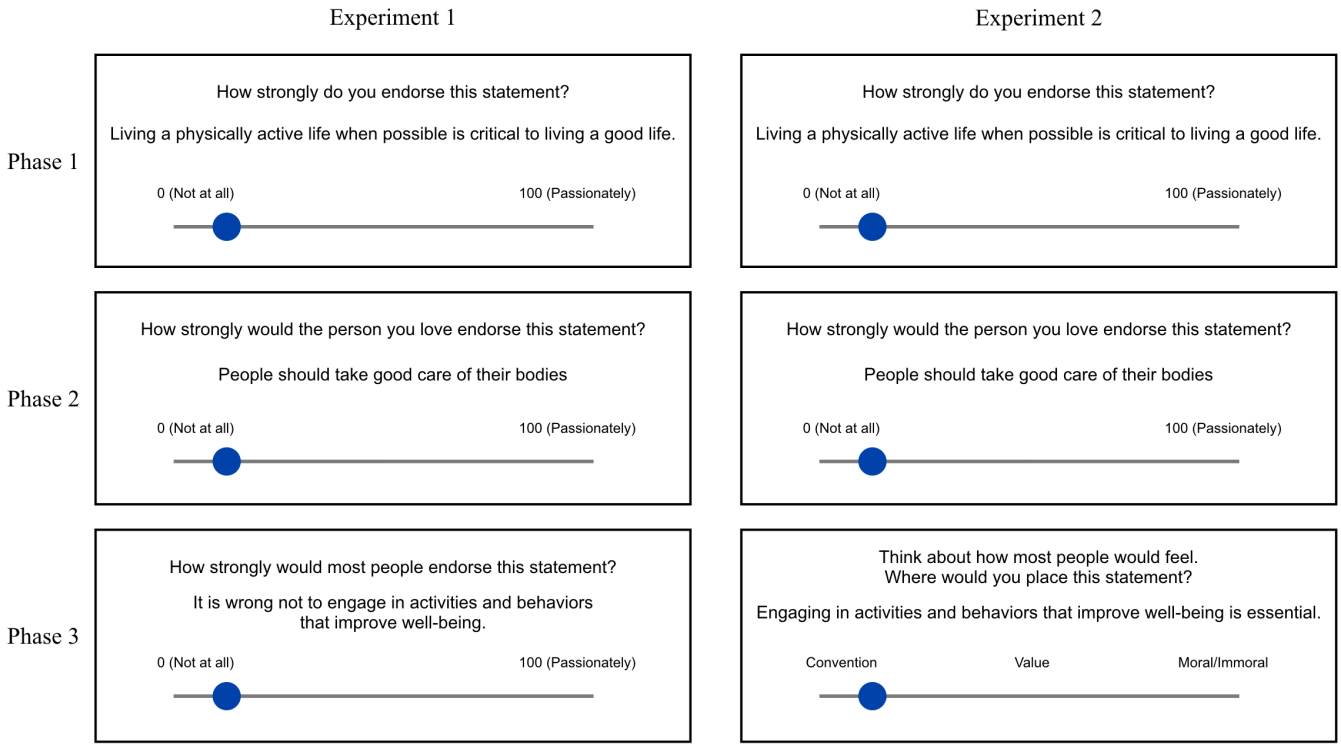

Figure 1. Examples from one trial from each phase of each experiment, drawn from the category of Athletics and Wellness. All examples are from the Close Other condition. Displays were identical in the Distant Other condition except that the word "love" in the Phase 2 question was changed to "know". Items were presented in a random order within each phase, and items were counterbalanced so that they appeared equally often in each phase.

Inclusion criteria.

Following Phase 3, participants were asked to report whom they were thinking about as a reference when they responded to the Phase 3 questions. If they did not answer that they were responding based on how most people would rate the statements (i.e., rather themselves or a

specific other person) they were excluded from the study. Twenty-seven participants were excluded for this reason. Additionally, at the end of the study participants were asked to re-enter the person or relationship they had thought of in Phase 2 to confirm it matched their initial response (suggesting they were actually thinking of this person throughout Phase 2). No participants were excluded for a mismatch between their initial and final answer. Finally, participants were excluded if they listed a seemingly close relationship (e.g., "Brother") in the Distant Other condition or a seemingly distant relationship ("co-worker) in the Close Other condition. Five participants were excluded for this reason. Finally, across the experiment, three check questions were inserted at random, one in each phase: twice participants were asked: *Tell us what the last statement you read was about* and once they were asked: *To show you are paying attention, please rate this statement to show you do not believe it is wrong*. Twenty participants were excluded for incorrect responses to the check questions.

Results

Our account predicted that if someone cared more about a behavior than you did, and you cared about that person, you would rate a failure to engage in that behavior as more impermissible than if you did not care about the other person.

To assess this we first identified categories in which participants' Phase 2 ratings (ratings of how much the other person cared) were higher than their Phase 1 ratings (ratings for themselves). Recall that participants rated different specific items in each phase, in different fixed, random orders, and were never asked to compare themselves directly with the other. To avoid imposing any arbitrary threshold on the difference scores, we used participants' raw scores; a "higher" rating refers to a numerical difference in the scores for the Phase 1 and Phase 2 item drawn from the same category. Individual items where participants did not provide a

response in any one of the three phases (n = 26 in total) were not included. In the Close Other condition, participants' rated the other person as caring more than themselves on 1071/2259 items (47.41%). In the Distant Other condition, participants rated the other person as caring more on 1030/2209 items (46.63%). There was no difference in the proportion of items where the other cared more than the participant between conditions, $X^2(1) = .24$, $p = .624$, $V = .008$. Additionally, participants' endorsement ratings did not differ by condition in either Phase 1 or Phase 2 (Phase 1: $t(4458.5) = -0.45$, $p = .655$, Close Other, mean: 64.46; Distant Other, mean = 64.85; CI for Difference of the means = [-2.06, 1.29]; Phase 2: $t(4459.4) = -0.06$, $p = .955$, Close Other, mean: 65.57; Distant Other, mean = 65.62; CI for Difference of the means = [-1.69, 1.60]).

We then compared the Phase 3 "wrongness" ratings between conditions, looking only at items from categories where the participants' Phase 1 and 2 scores indicated that the other person cared more about that kind of behavior than themselves. As predicted, participants in the Close Other condition judged failing to engage in these behaviors as more wrong than participants in the Distant Other condition: $t(2088.1) = 3.10$, $p = .002$ (Close Other, mean: 61.01; Distant Other, mean = 57.41; CI for Difference of the means = [1.32, 5.88]).

To see whether the results were driven by only a few category items or applied broadly, we looked at all participants who rated the other as caring more than themselves in a given category and averaged their Phase 3 scores for that item. Participants in the Close Other condition rated the behavior as more wrong than participants in the Distant Other condition in 16 of the 21 categories; that is, the mean Phase 3 rating was numerically higher in the Close Other condition than the Distant Other condition for a significant majority of the categories (binomial test, $p < .002$). See Table 1, Columns 2 and 3.

We can also ask whether the effect of Condition (Close versus Distant other) depends on the individual ratings in Phase 1 and Phase 2. We did this in two ways. First, we ran a regression analysis looking at the effect of Condition after including Phase 1 and Phase 2 ratings (as controls) in the model. Intuitively, items with higher Phase 1 and Phase 2 scores will have higher Phase 3 scores (i.e., because they are all drawn from the same behavioral construct). Nonetheless, in Experiment 1, Condition remained a significant predictor of Phase 3 ratings after controlling for the Phase 1 and Phase 2 ratings ($p = .011$). We can also ask whether the effect of moral alchemy held perhaps only for those behaviors rated relatively highly (i.e., those the participant valued greatly), or alternatively, only for those behaviors rated relatively low (i.e., those valued weakly). We performed a median split and found no interaction between high versus low Phase 1 or Phase 2 ratings and Condition on Phase 3 ratings (ps > .60). These results suggest that the effect of moral alchemy applies to both low and high-rated behaviors, and does not depend in any significant way on the initial rating given by participants either for themselves or for the other person.

Finally, to ensure that these effects were specific to cases where the participant believed the other person cared more, we compared the Phase 3 "wrongness" ratings between conditions, looking at cases where a participant had *not* rated the other person as caring more about that kind of behavior than themselves. As predicted, there was no difference between conditions in the Phase 3 ratings for these items: $t(2364.6) = 1.09$, $p = .277$ (Close Other, mean: 59.73; Distant Other, mean = 58.52; CI for Difference of the means = [-0.97, 3.40]).

Discussion

These results suggest that failing to engage in a behavior is perceived as "more wrong" by people in general when someone you care about cares more about the behavior than you do.

This is consistent with our account of moral alchemy, in which behaviors ranging from matters of negligible import to actual moral imperatives take on additional moral heft when a loved one cares more than you do. By hypothesis, this is because in caring but not distant relationships, failing to care sufficiently about what the other person cares about is a potential source of interpersonal harm, and thus in itself, a moral wrong; by contrast, if you already valued the behavior as much or more than your loved one, there is no such risk and therefore no additive moral concern.

It might of course be the case that thinking about a loved one leads people to value all positive behaviors more than they would otherwise (and thus view failing to act in accord with these behaviors as more than usually wrong). Our results however, suggest that this was not the case. Participants in the Close Other condition did not perceive failure to engage in all the positive behaviors as "more wrong" than those in the Distant Other condition; the effect was specific to those behaviors where the loved one cared more than the participant. This suggests that people were not motivated by concern for the loved one generally but that the act of caring about loved ones' priorities added moral urgency to the participants' own judgments.

Arguably, having just indicated that someone believed, relatively strongly, that a behavior was important in Phase 2, participants might have been more likely to provide higher ratings for comparable items in Phase 3. Relatedly, because people in Phase 3 were asked to respond on the basis of how "people in general" would respond, the effect of having just considered a specific other person's response might have anchored their responses or contributed to an availability heuristic, leading to the higher rating in Phase 3 (Tversky & Kahneman, 1973; 1974). Again however, our data suggest that this was not the case. The structure of the task was identical between conditions, and both the proportion of items about which people believed the

other person would feel more strongly, and the average endorsement ratings in Phase 1 and 2 were identical between conditions; however, participants' wrongness judgments were significantly stronger in the Close Other condition when they believed other person cared more.

It remains possible that the Phase 2 responses were nonetheless more salient (and therefore more likely to prime subsequent responses) in the Close Other than the Distant Other condition. However, we believe the experimental design makes a simple carry over effect very unlikely. In Phase 2, participants rated how passionately they cared about 21 different behaviors, each presented singly (so that participants could not track their responses across questions), and each from a different, unlabeled category. In Phase 3, participants responded to a different set of behaviors, in a different random order, and to a different question: not how much they cared about the behavior but how wrong it would be not to engage in it. Moreover, participants were never told that the behaviors were categorized in any way, and although items within a category were designed to relate to common themes (Davidov et al., 2008; Graham et al., 2009; Schwartz et al., 2001), there was relatively little overlap in the wording of the three items within a category. (See Appendix.) This design makes it very unlikely that participants could track, much less compare or be directly influenced, by their previous ratings within a category. Rather, we suggest that participants in the Close Other completed Phase 3 thinking about their loved one; when they came across a behavior that was more important to their loved one than themselves, concern for their loved one caused them to perceive failing to engage in that behavior as more impermissible.

Experiment 2

Experiment 1 looked at whether people believed third parties would see failing to engage in a behavior as more wrong when someone cared more than they did about a behavior, and they

cared about that person. In Experiment 2, we look at an even more direct measure of people's shift in moral concerns: shifts in the perceived significance of valued, positive behaviors.

As researchers have long noted, there are many beliefs and behaviors that are normative within a culture – e.g., habits of dress or manners – to which people broadly subscribe, but in which they are nonetheless not deeply invested; such social conventional norms can change by general agreement or by the will of an authority member (see Turiel, 1983; Nucci, 2001; though see Kelly, et al., 2007, Nichols & Fold-Bennett, 2003 for discussion). Other beliefs, which we will call values, may not be held as broadly, in that they vary from individual to individual, but are typically held more deeply: these include things like believing in the importance of academic achievement, creativity, athleticism, etc. (Hofstede, 1980; Kluckhohn, 1951; Rokeach, 1973; Schein, 1985; Schwartz, 1992; Smith & Schwartz, 1997; Williams, 1970). Both of these contrast with moral claims, which are held to have universal, objective force, independent of general agreement or authority (Cushman, 2008; Haidt, 2001; Kant., & Gregor, 1988; Mikhail, 2000; Rawls, 1971).

Our aim here is not to reify distinctions among these categories but to look at how caring about someone who cares about a behavior might affect the distinctions people make. That is, given that valued behaviors can exist "along a continuum of relative importance" (Rokeach, 1973, p. 5), we can ask whether moral alchemy shifts this relative importance. To do this, in Experiment 2 we replicated Phase 1 and 2 of Experiment 1 but changed the question in both the norming study and Phase 3 to ask participants where each behavior fell on a scale in which norms were the least important, values of intermediate importance, and morals of greatest importance. Again, in so doing we do not mean to imply that norms, values, and morals vary along only a single dimension; moral philosophers and psychologists have long discussed the

many ways these concepts diverge (e.g., Nucci, 1981). Here we merely make use of one dimension on which norms, values, and morals plausibly do vary continuously – relative importance – and ask whether the fact that someone you care about cares about a behavior more than you do causes you to raise your estimate of how third parties will view the significance of this behavior, effectively shifting it towards the moral end of this spectrum.

Methods

Participants.

Participants were recruited on Amazon Mechanical Turk and paid for their participation. One group of participants ($N = 38$, $m_{age} = 35.66$ years, 58% female) was used to norm the stimulus items. Another group of participants ($N = 305$) was recruited for the experimental conditions. Participants were excluded for 1) failure to answer attention check and comprehension questions correctly ($n = 31$) a mismatch between their initial and final response in identifying the target person ($n = 5$) (see Inclusion Criteria below for details), resulting in a final sample of 269 participants ($n = 131$ in Close Other condition, $m_{age} = 35.53$ years, 55% female; $n = 134$ in Distant Other condition, $m_{age} = 34.44$ years, 54% female).

Materials.

The same items from Experiment 1 were used. In all phases of the experiment, items were phrased as in Appendix, Part 1. Items were counterbalanced as in Experiment 1.

Procedure.

Stimulus norming.

Participants were also tested online. They were shown the sliding scale as it appears in Figure 1, Experiment 2, Phase 3 that they could manipulate with their computer mouse. They

were told that they would be given a series of statements and asked to judge how much they agreed with each statement:

There are some beliefs people hold, and expect others to hold, but don't feel too strongly about (like believing the napkin belongs on the left side of the plate). We will call these conventions. For other kinds of beliefs, people differ widely. Some people care a great deal and others not at all (like believing that theater is important). We will call these values. Other beliefs people feel strongly about (like believing stealing is wrong) and expect others to feel strongly about. We will call these morals.

Conventions, values, and morals differ in many ways. But here we will ask you to consider how much most people might be expected to care about the following statements. You will be asked to rank each statement on a continuous scale from: ***Convention:*** *Something that on average, people don't feel strongly about even if they expect other people to share their expectations.* ***Value****: Something that some people might care deeply about and others not all. Thus on average, people feel more strongly about these beliefs than conventions but less strongly than they do about morals.* ***Moral/Immoral****: Something that almost everyone feels strongly about. People may disagree about what they feel, but most people care a great deal about the content of these statements. Although these statements are about conventions, values, and what is moral/immoral, we are not asking how much you hold each belief. So when responding using sliding scale, you do not need to consider whether you personally agree with the statement in order to answer. Rather think about how most people would feel and use the slider to rank the statements accordingly.*

Participants were shown five pictures of scale, each with the pointer at different locations to indicate they could place the pointer at any point on the scale.

Again, the norming study was used to establish 1) that participants understood the task and used the sliding scale appropriately with respect to common sense judgments; 2) that there was variability both within and across items in participants' responses, and 3) that the items were easy to read. All of these results were confirmed. Overall, the statements received a mean rating of 42.03 (20.92); range 3.84-95.75. Participants used the scale in a meaningful way. For instance a statement that intuitively reflected a social convention ("*People should use proper etiquette …*"), a personal value ("*Being able to provide financially for yourself and others is one of the most important abilities*"), and a moral concern ("*Justice and equality are important …*"), received mean ratings of 13.11, 45.11, and 86.11 respectively. The average standard deviation for individual participants' responses within items was 21.45 points, suggesting again that there was sufficient variability in individuals' judgments to test our hypothesis. Overall, the readability of the items were rated an average of 1.52 (.29), or between *Very Easy* and *Easy.*

Experimental design: Phases 1-3.

Phases 1 and 2 were identical to Experiment 1. Phase 3 was identical to the norming study except that participants were not asked to rate the readability of each statement.

Results

Experiment 2 tested the prediction that if someone cared more about a behavior than you did, and you cared about that person, you would place the behavior closer to the moral end of the scale than if you did not care about the other person, or the other person did not care about the behavior.

Table 1. Mean Phase 3 ratings by condition for each category where the other person cared more than the participant. Categories are listed in order of their mean ratings from the norming studies

for Experiment 1 and Experiment 2 respectively. The numerically higher mean rating is boldfaced.

Categories	Exp. 1 Close Other	Exp. 1 Distant Other	Categories	Exp. 2 Close Other	Exp. 2 Distant Other
1) Harm & Welfare	**87.02**	76.51	1) Harm & Welfare	**87.98**	86.51
2) Fairness	74.26	**74.65**	2) Fairness	**80.43**	75.98
3) Social Behavior	**73.91**	69.37	3) Loyalty	**65.69**	64.52
4) Universalism & Environmentalism	**76.00**	61.86	4) Purity & Sanctity	**63.71**	62.44
5) Authority	**73.76**	68.60	5) Authority	57.77	**58.64**
6) Athletics & Welfare	**65.23**	57.05	6) Religion	**54.79**	50.26
7) Loyalty/In-Group	**72.89**	66.07	7) Tradition	48.80	**51.51**
8) Regulations	**64.14**	56.56	8) Achievement & Self-Direction	**49.88**	46.69
9) Purity & Sanctity	**66.83**	57.33	9) Universalism & Environmentalism	**69.23**	57.23
10) Scholarship	**61.42**	58.24	10) Scholarship	43.63	**45.95**
11) Politeness	**61.25**	54.63	11) Financial Power	**39.95**	37.90
12) Achievement & Self-Direction	**62.33**	55.52	12) Social Behavior	**44.35**	43.73
13) Enjoyment	54.10	**54.28**	13) Openness	**41.61**	33.50
14) Dress	**48.63**	**45.91**	14) Athletics & Wellness	**48.22**	45.53
15) Openness	51.12	**52.13**	15) Regulations	**37.15**	33.17
16) Tradition	**56.15**	54.02	16) Curiosity & Creativity	**41.92**	35.92
17) Art & Aesthetics	**43.73**	42.96	17) Politeness	**40.32**	34.34
18) Religion	**52.39**	51.07	18) Enjoyment	**44.77**	39.82
19) Organization & Neatness	**50.80**	48.09	19) Art & Aesthetics	**34.29**	32.46
20) Financial Power	50.81	**56.98**	20) Organization & Neatness	15.05	**19.70**
21) Curiosity & Creativity	43.52	**46.72**	21) Dress	16.33	**20.20**

We looked first at categories where the participant believed the other person cared more, corresponding to a higher rating in Phase 2 than Phase 1 for items drawn from a single category. As in Experiment 1, we did not impose any threshold on the difference scores. We used participants' raw scores: a "higher" rating refers to a numerical difference in the scores for the Phase 1 and Phase 2 item drawn from the same category. Individual items where participants did not provide a response in any one of the three phases (n = 64 in total) were not included. In the Close Other condition, participants rated the other person as caring more on 1203/2734 (44.00%); in the Distant Other condition, there were 1278/2768 such items (46.17%). There was no difference in the proportion of items in which participants rated the close other as caring more, although there was a slight trend towards believing the other cared more on more items in the Distant Other condition ($X^2(1) = 2.53$, $p = .112$, $V = .022$). Additionally, participants' mean endorsement ratings did not differ by condition in either Phase 1 or Phase 2 (Phase 1: $t(5496.9) = .77$, $p = .444$, Close Other, mean: 68.18; Distant Other, mean = 67.63; CI for Difference of the means = [-.86, 1.96]; Phase 2: $t(5338.5) = -1.69$, $p = .090$, Close Other, mean: 66.34; Distant Other, mean = 67.55; CI for Difference of the means = [-2.63, 0.19].

We then looked at how people rated items in Phase 3, looking at the categories where the participants' Phase 1 and 2 scores indicated that the other person cared more about that kind of behavior than themselves. As predicted, participants in the Close Other condition placed these behaviors further towards the moral end of the scale than participants in the Distant Other condition: $t(2470.1) = 2.09$, $p = .037$ (Close Other, mean: 48.09; Distant Other, mean = 45.63; CI for Difference of the means = [.15, 4.77]).

As in Experiment 1, we also looked at whether the difference between Close and Distant Other conditions applied broadly across the categories. Participants in the Close Other condition rated items from categories in which the other cared more further towards the moral end of the scale than participants in the Distant Other condition on 16 of the 21 categories; that is, the mean Phase 3 rating was numerically higher in the Close Other condition than the Distant Other condition for a significant majority of the categories (binomial test, $p < .002$). See Table 1.

Also as in Experiment 1, we ran a regression analysis looking at the effect of Condition after including the other variables (as controls) in the model. Recall that items with higher Phase 1 and Phase 2 scores will have higher Phase 3 scores because they are different behaviors drawn from the same category. Nonetheless, in Experiment 2, there was a trend for Condition to predict Phase 3 ratings even after controlling for the Phase 1 and Phase 2 ratings ($p = .069$). A median split looking at whether there was an interaction between high versus low Phase 1 or Phase 2 ratings and Condition on Phase 3 ratings found no interaction (ps > .20). These results suggest that the effect of moral alchemy is robust to the individual Phase 1 and Phase 2 ratings.

Finally, to ensure that these effects were specific to cases where the participant believed the other person cared more, we compared the Phase 3 ratings between conditions, looking at cases where participants had *not* rated the other person as caring more about that kind of behavior than themselves. Again, as predicted, there was no difference between conditions in the Phase 3 ratings for these items: $t(3014.1)=.17$, $p=.866$ (Close Other, mean: 48.40; Distant Other, mean = 48.22; CI for Difference of the means = [-1.97, 2.34]).

Discussion

Experiment 1 suggested that failing to engage in a behavior is perceived as "more wrong" by third parties when someone you care about cares more about the behavior than you do.

Experiment 2 suggests that the behavior may also be perceived as "more moral". To the degree that positive behaviors exist on a continuum of importance, with conventions regarded as relatively unimportant, values as moderately important (insofar as their importance varies from person to person), and morals as extremely important, believing that a loved one cares more than you about a behavior elevates the significance of the behavior, making conventions somewhat more like values and values somewhat more like morals.

The results further suggest that this was not due to a general elevation of the status of positive behaviors in the context of thinking about a loved one. Relative to participants in the Distant Other condition, participants' ratings in the Close Other condition were only higher when they believed the loved one cared more about that kind of behavior than they did themselves. This is consistent with the idea that concern for the interests of close others makes us treat some behaviors more seriously than we otherwise would because failing to so value them risks interpersonal harm.

As in Experiment 1, we believe the results are unlikely to be due to participants' Phase 2 estimates directly influencing their Phase 3 estimates. Both the proportion of behaviors identified as more valuable to the other than the self, and the degree to which behaviors were valued, were similar between conditions in Phase 1 and Phase 2. However, despite comparable responses in the preceding phases (and thus comparable opportunities to carry over the earlier responses, or to treat the earlier estimates as representative) participants shifted behaviors further towards the moral end of the scale in the Close than Distant Other condition of Phase 3. Although thinking about a close relationship might have made the earlier ratings more salient, as in Experiment 1, the design made it unlikely that participants could track (and thus be directly primed by) their previous ratings. Rather, we believe that in the Close Other condition, concern

for the loved one added moral importance to behaviors perceived as more important to their close other than the self, leading participants to elevate the moral significance of these behaviors to third parties.

General discussion

The current study provided a preliminary test of the idea that caring relationships can lead to a kind of moral alchemy. We proposed that when we believe a loved one cares about a behavior more than we do ourselves, the moral imperative to care about the loved one's interests raises the perceived value of that behavior to third parties, such that violations of the behavior are seen as more wrong (Experiment 1) and the behavior itself as more moral (Experiment 2) than when we think about values outside of the context of an intimate relationship. The current study confirmed these predictions.

We do not want to overstate the results here. In both Experiments 1 and 2, the differences between participants' mean ratings of the moral status of the behaviors in the two conditions were small (3.60 points in Experiment 1 and 2.46 points in Experiment 2). The small change in ratings may seem insufficient to warrant the term "moral alchemy" insofar as alchemy implies a transformation from one kind of thing into another (e.g., a convention or value into a moral). However, for some individual items, the average difference between the close and distant other conditions were fairly striking (e.g., more than ten points for values of Universalism and Environmentalism in both Experiment 1 and 2), suggesting that at least some behaviors may indeed have transformed from having a status more like values to a status more like morals. Given that participants were asked merely to imagine a close other, in the context of an online Internet survey, with no personalized content, and no real world ramifications, we find it striking

that there was a relative change in people's evaluations of behavior consistent with "moralizing" behaviors that might, if undervalued, lead to interpersonal harm or violations of care.

Note that participants were randomly assigned to the Close versus Distant Other condition and the results are consistent with the causal claim that consideration of the values of close others elevates the degree to which those values are seen as moral. As predicted, we also found that these results held only when the participants believed the other person cared more than they did (and not when the participants believed the other person cared less). However, the others' values were reported, not manipulated. Thus although it is possible that thinking about a loved one caring more led participants to raise the moral status of the behavior, other interpretations are possible. For instance, those who believed that people in general saw a behavior as more moral than they did may have also believed their close other valued it more; alternatively, a common cause might have led to higher ratings for both the close other and people in general (e.g., thoughts of caring might have led participants to exaggerate the difference between their own and others' estimate of some behaviors). Future research might manipulate the information provided about the loved one's values to disambiguate these interpretations.

We have already discussed the respects in which we believe these results cannot be explained by simple carry over effects, or by a general enhancement of the importance of positive behaviors in the context of loving relationships. For similar reasons, the results can also not be explained as a kind of "chameleon effect" in which people automatically and unconsciously take on others' attributes or behaviors others (Chartrand & Bargh, 1999). While it may be generally valuable for close kin and members of in-group to align with each other's values (see Cialdini & Goldstein, 2004 for review), in this context, participants' tendency to rate

wrongness or moral significance in alignment with the perceived values of the close other was uni-directional. Participants considering close (versus distant) others rated behaviors as more moral when the other person cared more, but there was no effect when the other person cared less. These results are consistent with the idea that participants' took their loved one's utilities into account only when failure to do so might risk hurting the other and was therefore an independent moral concern.

Note that although participants treated behaviors valued more by loved ones than themselves as more morally important to the world at large, this does not mean that participants necessarily changed their personal estimate of the behaviors. Our experiment did not address this directly (because endorsements for the self always preceded estimates for the loved one). However, on our account, it is possible to remain relatively indifferent to a behavior – say the importance of appreciating art – and nonetheless (if a loved one cares deeply about it) be more likely to feel that third parties will perceive it as (somewhat) wrong not to appreciate art, and elevate art appreciation to something more like a moral commitment than a value. The point of moral alchemy is precisely to explain everyday moral dilemmas where we experience a tension between what we *actually* believe and the need to consider the values of close others. We explicitly do *not* propose that we therefore change our minds and come to view (like them) that the value is important in its own right. Rather we propose that (relative to cases where these views as held by strangers, or where our close other cares less than we do) we are more likely to recognize that violations of these behaviors could do harm, and thus see these values as mattering morally to third parties. Indeed, we have suggested that part of the effect of moral alchemy is to expand the realm of moral concerns and make us value things for the sake of others that we would not value as highly on our own. Thus our particular dependent measures focused

not on how a close other valuing the behavior changed how participants themselves rated the behavior but instead focused on how participants believed people in general would value the behavior.

In line with most work on moral reasoning, we also looked here at moral judgment rather than moral behavior (and there is evidence that hypothetical moral judgments do not necessarily align with behavior; see e.g., Crockett et al., 2014). However, in the case of moral alchemy it is especially true that shifts in moral judgment may not be reflected in all morally relevant behaviors. If for instance, someone does not care greatly about some behaviors (academic achievement, homosexuality, etc.) except insofar as her loved one does, she might be disposed both to elevate their status as moral behaviors and yet (outside the realm of potential harm to her loved one) not be more inclined to enact or refrain from such behaviors, or punish third party transgressors. Future research might look at whether the current results converge both with more overt behavioral measures (e.g., how people actually behave in moral decision-making contexts) and more implicit measures that do not rely on explicit judgments.

Similarly, more work is necessary to understand why people shift their judgments in close relationships but not more distant ones. Future work might see whether the shifts in moral judgment seen here correlate with particular measures of closeness in the social psychology literature. For instance, people might be more likely to shift their judgments in relationships which include high degrees of self-disclosure and "including the other in the self" (Aron, Aron, & Smollan, 1992) or they might be more subject to moral alchemy in relationships where partners routinely generate constructive responses to potentially destructive behavior and manifest high interpersonal commitment (Rusbult, Verette, Whitney, Slovik, & Lipkus, 1991). Although here we have focused largely on benign or positive aspects of moral alchemy, it may

have a darker side as well. We have emphasized the tendency to adjust our values in the direction of elevating the moral status of behaviors a loved one cares about; however, we may also devalue things important to ourselves to the extent that what we care about might hurt those we love. Additionally, as long as the obligations of interpersonal care are unevenly distributed in society (see e.g., Held, 1993; Tronto, 1993), women may be more likely than men to be torn between their own values and alchemical ones – values held for their own sake and those held because the moral responsibilities associated with care invest them with value. Finally, we note that the ability to recursively value the values of others (Kleiman-Weiner & Tenenbaum this issue; Jara-Ettinger, et al., 2015; Jara-Ettinger, et al., 2015; 2016; Ullman, et al., 2010; Yoshida, et al., 2008) is a necessary condition for moral behavior, but not a sufficient one. If someone gains utility from a destructive behavior and we support them, we may have promoted the good of another but we have failed to promote "the good" in any larger sense. In this sense, the ability to recursively link one's own utilities to another's falls short of philosophical accounts of recursive value which ground out in "intrinsic goods" (and recursively, assert that actions which promote intrinsic goods are good, and that loving those actions which promote those goods is good; Hurka, 2003). However, as noted, real world moral dilemmas often center on precisely those cases where we worry that we cannot both support the other and support his values. In such cases, we may try to "hate the sin and love the sinner" but we cannot avoid concerns of harm to the degree that the "sinner" feels hurt by that kind of love.

We have asserted that failing to endorse the values of close others risks hurting their feelings (and is therefore a legitimate source of moral concern). However we have elided the question of when and why value differences should be painful. That they are not always so is clear; many of us have delighted in debate and intellectual sparring with intimate friends and

family. It may be that such experiences are delightful because they attest to attachments so secure that even conflict is "safe". However in many other contexts, disagreement, especially but not exclusively over moral matters, causes social discomfort and threatens the integrity of groups (Aronson, 1999; Festinger, 1957; Heider, 1958; Kennedy & Pronin, 2014; Klucharev, Hytönen, Rijpkema, Smidts, & Fernandez, 2009; Matz & Wood, 2005; Skitka et al., 2005; Stone & Cooper, 2001, 2003; Wainryb, Shaw, Laupa, & Smith, 2001; Wainryb, Shaw, & Maianu, 1998). In this respect moral alchemy may amount to "moral pragmatism": attributing more importance to matters when someone close to you is more invested than you are may be not just a caring thing to do, but a prudent one.

Finally, although here we have tested our account of moral alchemy with adults, future work might look at whether the findings here extend to young children and whether it leads not just to transient shifts in moral judgments but to more enduring behavioral changes. To the degree that it does, the phenomenon is potentially most relevant to learning a system of values in early childhood. In adding moral weight to the issues their parents care about most, moral alchemy may support children's internalization of values (through "referencing the absent parent"; Thompson, Meyer, & Mcginley, 2006; see also Emde, Biringen, Clymna, & Oppenheim, 1991; Emde & Buchsbaum, 1990) and contribute to the fidelity of cultural transmission. Over time however, the tension between individuals' beliefs and the responsibility of care could also contribute to social change. In inducing us to adjust even our most intransigent beliefs out of concern for others, moral alchemy may help transform our collective moral imagination.

Appendix

Part 1. Items for Experiment 1, Phases 1-2 and for Experiment 2, norming study, and Phases 1-3.

Achievement & Self-Direction
Showing what one is capable of and reaching one's full potential is crucial.
People should make their own decisions and forge their own paths.
Being successful in one's chosen career is of utmost importance.

Art & Aesthetics:
People should be aware of the look and feel of their environment and aim to make their surroundings more pleasing to the eye.
It is essential to attend to the visual qualities of one's environment.
Making as well as appreciating art is of utmost importance.

Athletics & Wellness
Living a physically active life when possible is critical to living a good life.
People should take good care of their bodies
Engaging in activities and behaviors that improve wellbeing is essential.

Authority
Even when soldiers disagree with their commanding officer's orders, they should obey anyway because that is their duty.
It is wrong when someone purposefully doesn't fulfill the duties of their role.
Respect for authority is something all children need to learn.

Creativity & Curiosity
Finding out how and why something works is of central importance.
Thinking up new ideas is an important endeavor and everyone should attempt to think of new ideas.
Developing insightful ways of doing things makes the world a better place, and people should look to add imaginative approaches to their lives.

Dress
People should wear clean clothes.
People should keep their hair neat.
Wearing clothes that are appropriate for the season is important.

Enjoyment
People should try to experience "the good things" in life whenever possible.
Once should seek every chance they have to have fun.
It is essential to look for pleasure in one's experiences.

Fairness
Justice and equality are important considerations.

When voting on laws, one important principle should be ensuring that everyone is treated fairly.
An action can be judged as right or wrong depending on whether someone was treat unjustly.

Financial Power
Being able to provide financially for yourself and others is one of the most important abilities.
It is important to be well off, and have enough money to buy things you and your family desire.
Having wealth is something that people should aspire to.

Harm & Welfare
Compassion for those who are suffering is a crucial virtue.
It is wrong to cause emotional suffering.
It's not right to kill a human being.

Loyalty/In-group
People should not betray their group.
People should be loyal to their family members, even when they have done something wrong.
People should ensure the well-being of those close to them.

Openness
People should always look for new things to do. It is essential to do lots of different things during a lifetime.
Taking risks and having an exciting life is one of the most important factors in determining whether one is living well.
It is important to consider ideas that may be initially unfamiliar.

Organization & Neatness
People should be orderly in their personal space, keeping their homes and offices neat.
People should follow certain rules to be neat, for example placing forks to the left of the plate when setting a table.
People should write neatly, and in school, students should check their work over carefully before completing it.

Politeness
People should use proper etiquette, such as eating foods like salad and pasta with utensils.
People should be courteous, holding doors open for others and letting elderly and young people go first.
People should avoid using profane language.

Purity
Some acts wrong on the grounds that they are unnatural.
People should follow standards of purity and decency.
People should not do things that are disgusting, even if no one is harmed.

Regulations
People should follow the rules when playing new games and activities.
People should raise their hands and wait to be called on before speaking in classes or meetings.

One should follow regulations even when there are no other participants, like using turn signals when changing lanes, even when there are no other cars on the road and no police nearby.

Religion
Being in touch with sentiments beyond the material is of utmost importance.
Practicing a faith adds value to people's lives.
Religion is an important aspect of one's life.

Social Behavior
A respect for personal space is of utmost importance. One shouldn't hug someone or a stranger without asking.
People should be careful to phrase their comments to others in a way that doesn't offend.
One shouldn't ask delicate or intimate questions of strangers.

Scholarship
One should seek to learn more, for example by studying and reading.
Working hard in school is something all children should strive to do.
People should get as much education as possible.

Tradition
People should try to follow the rituals practiced by their families.
Following the customs of your family is an important way of honoring them.
Modesty is a central virtue.

Universalism & Environmentalism
People should listen to people who are different from themselves to gain perspective.
People should care for the earth. Looking after the environment is essential.
It is essential to protect wildlife habitats.

Part 2. Items from Experiment 1, norming study and Phase 3.

Achievement & Self-Direction
It is wrong not to show what one is capable of and not reach one's full potential.
It is wrong for people not to make their own decisions, and not to forge their own paths.
It is wrong not to be successful in one's chosen career.

Art & Aesthetics:
It is wrong for people not to be aware of the look and feel of their environment, and not to aim to make their surroundings more pleasing to the eye.
It is wrong not to attend to the visual qualities of one's environment.
It is wrong not to make, as well as appreciate, art.

Athletics & Wellness
It is wrong not to live a physically active life, when possible.
It is wrong for people not to take good care of their bodies.

It is wrong not to engage in activities and behaviors that improve wellbeing.

Authority
It is wrong for soldiers to disagree with their commanding officer's orders, since they should obey anyway because that is their duty.
It is wrong when someone purposefully doesn't fulfill the duties of their role.
It is when wrong when children don't learn to respect authority.

Creativity & Curiosity
It is wrong not to find out how and why something works.
It is wrong not to attempt to think of new ideas.
It is wrong not to develop insightful ways of doing things and for people not to look to add imaginative approaches to their lives.

Dress
It is wrong for people not to wear clean clothes.
It is wrong for people not to keep their hair neat.
It is wrong not to wear clothes that are appropriate for the season.

Enjoyment
It is wrong for people not to try to experience "the good things" in life whenever possible.
It is wrong for people not to seek every chance they have to have fun.
It is wrong not to look for pleasure in one's experiences.

Fairness
It is wrong for people not to treat justice and equality as important considerations.
It is wrong not to try to ensure that everyone is treated fairly when voting on laws.
It is wrong to treat someone unjustly.

Financial Power
It is wrong not to be able to provide financially for yourself and others.
It is wrong for people not to be well off, and not to have enough money to buy things you and your family desire.
It is wrong when people do not aspire to have wealth.

Harm & Welfare
It is wrong not to show compassion for those who are suffering.
It is wrong to cause emotional suffering.
It is wrong to kill a human being.

Loyalty/In-group
It is wrong for people to betray their group.
It is wrong for people not to be loyal to their family members when they have done something wrong.
It is wrong for people not ensure the well-being of those close to them.

Openness
It is wrong for people not to continually look for new things to do during a lifetime.
It is wrong not to take risks and not to have an exciting life.
It is wrong not to consider ideas that may be initially unfamiliar.

Organization & Neatness
It is wrong for people not to be orderly in their personal space by keeping their homes and offices neat.
It is wrong for people not to follow certain rules for being neat, for example not placing forks to the left of the plate when setting a table.
It is wrong for people not to write neatly, and in school, for students not to check their work over carefully before completing it.

Politeness
It is wrong for people not to use proper etiquette, for example eating foods like salad and pasta without utensils.
It is wrong for people not to be courteous, for example not holding doors open for others and letting elderly and young people go first.
It is wrong for people to use profane language.

Purity
It is wrong to do some acts because they are unnatural.
It is wrong for people not to follow standards of purity and decency.
It is wrong for people do things that are disgusting, even if no one is harmed.

Regulations
It is wrong for people not to follow the rules when playing new games and activities.
It is wrong when people do not raise their hands and wait to be called on before speaking in classes or meetings.
It is wrong not to follow regulations when there are no other participants, like not using turn signals when changing lanes, even when there are no other cars on the road and no police nearby.
It is wrong not to be in touch with sentiments beyond the material.

Religion
It is wrong not to be in touch with sentiments beyond the material.
It is wrong not to practice a faith, because faith could add value to people’s lives.
It is wrong for religion not to be an important aspect of one's life.

Social Behavior
It is wrong not to respect others' personal space, for example by hugging someone or a stranger without asking.
It is wrong when people are not careful to phrase their comments to others in a way that doesn’t offend.
It is wrong for one to ask delicate or intimate questions of strangers.

Scholarship
It is wrong not to seek to learn more, for example by studying and reading.
It is wrong for children not to work hard in school.
It is wrong for people not to get as much education as possible.

Tradition
It is wrong for people not to try to follow the rituals practiced by their families.
It is wrong not to follow the customs of your family to honor them.
It is wrong not to have modesty as a central virtue.

Universalism & Environmentalism
It is wrong for people not to listen to people who are different from themselves to gain perspective.
It is wrong for people not to care for the earth.
It is wrong not to protect wildlife habitats.

Acknowledgements

We thank the participants in these studies, and Rebecca Saxe and Josh Tenenbaum for discussion. This material is based upon work supported by the Center for Brains, Minds, and Machines (CBMM), funded by NSF STC awards CCF-1231216.

References

Aknin, L. B., Hamlin, J. K., & Dunn, E. W. (2012). Giving leads to happiness in young children. *PLoS One*, *7*, e39211.

Anderson, E. S. (1999). What Is the Point of Equality? *Ethics*, *109*, 287-337.

Aristotle. & Sachs, J. (2002). Nicomachean ethics. Newbury, MA: Focus Pub./R. Pullins.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*, 596.

Aronson, E. (1999). Dissonance, hypocrisy, and the self-concept. *Readings about the Social Animal*, 219-236.

Bandura, A., & McDonald, F. J. (1963). Influence of social reinforcement and the behavior of models in shaping children's moral judgment. *The Journal of Abnormal and Social Psychology*, *67*, 274.

Barni, D., Knafo, A., Ben-Arieh, A., & Haj-Yahia, M. M. (2014). Parent–child value similarity across and within cultures. *Journal of Cross-Cultural Psychology*, *45*, 853-867.

Barragan, R. C., & Dweck, C. S. (2014). Rethinking natural altruism: Simple reciprocal interactions trigger children's benevolence. *Proceedings of the National Academy of Sciences*, *111*, 17071-17074.

Baumrind, D. (1986). Sex differences in moral reasoning: Response to Walker's (1984) conclusion that there are none. *Child Development*, *57,* 511-521.

Benenson, J. F., & Dweck, C. S. (1986). The development of trait explanations and self-evaluations in the academic and social domains. *Child Development*, *57*, 1179-1187.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, *57*, 1-29.

Blair, R. J. R. (2009). Neurocognitive systems involved in moral reasoning. In *The Moral Brain* (pp. 87-107). Springer Netherlands.

Blair, J., Marsh, A. A., Finger, E., Blair, K. S., & Luo, J. (2006). Neurocognitive systems involved in morality. *Philosophical Explorations*, *9*, 13-27.

Boer, D., & Fischer, R. (2013). How and when do personal values guide our attitudes and sociality? Explaining cross-cultural variability in attitude–value linkages. *Psychological Bulletin*, *139*, 1113.

Brandt, A. M., & Rozin, P. (2013). *Morality and health*. New York, NY Routledge.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*, 307.

Brody, G. H., & Henderson, R. W. (1977). Effects of multiple model variations and rationale provision on the moral judgments and explanations of young children. *Child Development*, 48, 1117-1120.

Brody, G. H., & Shaffer, D. R. (1982). Contributions of parents and peers to children's moral socialization. *Developmental Review*, *2*, 31-75.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological inquiry*, *6*, 1-30.

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domain. *Social Development*, *6*, 53-75.

Campbell, R., & Kumar, V. (2013). Pragmatic naturalism and moral objectivity. *Analysis*, *73*(3), 446-455.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591-621.

Cikara, M., Farnsworth, R. A., Harris, L. T., & Fiske, S. T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social cognitive and affective neuroscience*, *5*, 404-413.

Cowan, P. A., Longer, J., Heavenrich, J., & Nathanson, M. (1969). Social learning and Piaget's cognitive theory of moral development. *Journal of Personality and Social Psychology*, *11*, 261.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*, 17433-17438.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, *111*, 17320-17325.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353-380.

Cushman, F. (2015). From moral concern to moral constraint. *Current Opinion in Behavioral Sciences*, *3*, 58-62.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological Science*, *17*, 1082-1089.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt.

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in the adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, *72*, 420-445.

Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. *Cerebral Cortex*, *22*, 209-220.

Dorr, D., & Fey, S. (1974). Relative power of symbolic adult and peer models in the modification of children's moral choice behavior. *Journal of Personality and Social Psychology*, *29*, 335.

Eisenberg, N., & Valiente, C. (2002). Parenting and children's prosocial and moral development. *Handbook of parenting*, *5*, 111-142.

Emde, R. N., Biringen, Z., Clyman, R. B., & Oppenheim, D. (1991). The moral self of infancy: Affective core and procedural knowledge. *Developmental Review*, *11*(3), 251-270.

Emde, R. N. & Buchsbaum, H. K. (1990). "Did you hear my mommy?" Autonomy with connectedness in moral self emergence. *Development of the self through the transition*, 35-60.

Eskine, K. J. (2013). Wholesome foods and wholesome morals? Organic foods reduce prosocial behavior and harshen moral judgments. *Social Psychological and Personality Science*, *4*, 251-254.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford university press.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5-15.

Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior.* Cambridge, MA: Belknap Press of Harvard University Press.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029.

Gray, K., & Ward, A. F. (2011). The harm hypothesis: Perceived harm unifies morality. *Manuscript submitted for publication*.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105-2108.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work?. *Trends in Cognitive Sciences*, *6*, 517-523.

Grusec, J. E., Goodnow, J. J., & Kuczynski, L. (2000). New directions in analyses of parenting contributions to children's acquisition of values. *Child Development*, *71*, 205-211.

Haan, N., Smith, M. B., & Block, J. (1968). Moral reasoning of young adults: political-social behavior, family background, and personality correlates. *Journal of Personality and Social Psychology*, *10*, 183.

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

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998-1002.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98-116.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The Innate Mind*, *3*, 367-392.

Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Unpublished manuscript, University of Virginia*.

Haidt, J., & Hersh, M. A. (2001). Sexual Morality: The Cultures and Emotions of Conservatives and Liberals. *Journal of Applied Social Psychology*, *31*, 191-221.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613.

Hamlin, J. K. (2013). Moral Judgment and Action in Preverbal Infants and Toddlers Evidence for an Innate Moral Core. *Current Directions in Psychological Science*, *22*, 186-193.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*, 30-39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557-559.

Hamlin, J.K, Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science, 13,* 923-929.

Harding, S. G. (1987). *Feminism and methodology: Social science issues*. Indiana University Press.

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological review*, *102*, 458.

Harris, J. R. (2011). *The nurture assumption: Why children turn out the way they do*. New York, NY: Simon and Schuster.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Held, V. (1993). *Feminist morality: Transforming culture, society, and politics*. Chicago, IL: University of Chicago Press.

Heller, K. A., & Berndt, T. J. (1981). Developmental changes in the formation and organization of personality attributions. *Child Development*, *52,* 683-691.

Helwig, C.C., & Turiel, E. (2003). Children's social and moral reasoning. In P.K. Smith & C.H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 475-490). Oxford, UK: Blackwell.

Heyman, G. D., Dweck, C. S., & Cain, K. M. (1992). Young children's vulnerability to self-blame and helplessness: Relationship to beliefs about goodness. *Child Development*, *63,* 401-415.

Hoffman, M. L. (1970). Moral development. In P. Musen (Ed.), *Carmichael's Manual of Child Psychology* (Vol. 2). New York, NY: Wiley and Sons.

Hoffman, M. L. (1975). Altruistic behavior and the parent-child relationship. *Journal of Personality and Social Psychology*, *31*, 937.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*, 1340-1343.

Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad?. *Organizational Dynamics*, *9*, 42-63.

Hurka, T. (2003). *Virtue, vice, and value.* Oxford, UK: Oxford University Press.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14-23.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends in Cognitive Sciences*, *20*, 589-604.

Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not So Innocent Toddlers' Inferences About Costs and Culpability. *Psychological science*, *26*, 633-640.

Joyce, R. (2006). *The evolution of morality.* Cambridge, MA: MIT Press.

Kant, I., & Gregor, M. J. (1988). *Groundwork for the metaphysics of morals.* Cambridge, UK: Cambridge University Press.

Katz, L. D. (2000). *Evolutionary origins of morality: Cross disciplinary perspectives.* Thorverton, UK: Imprint Academic.

Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, *22*, 117-131.

Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*, 833-848.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*, 12577-12580.

Kleinman-Weiner, M., & Tenenbaum, J.B. (submitted). Learning a commonsense theory of morality.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, *61*, 140-151.

Kluckhohn, F. (1951). Values and Value-Orientation in the Theory of Action: An Exploration in Definition and Classification in Parsons T, Shils EA (Eds.) *Towards a general theory of action.* Cambridge, MA: Harvard University Press.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*, 190-194.

Knobe, J. (2004). Folk Psychology and Folk Morality: Response to Critics.

Kochanska, G. (1997). Multiple pathways to conscience for children with different temperaments: from toddlerhood to age 5. *Developmental Psychology*, *33*, 228.

Kohlberg, L. (1969). Stage and sequence: The cognitive-develop- mental approach to socialization. In D.A. Goslin (Ed.), Handbook of socialization theory and research (pp. 151–235). New York, NY: Academic Press.

Kuczynski, L. (1984). Socialization goals and mother–child interaction: Strategies for long-term and short-term compliance. *Developmental Psychology*, *20* 1061.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*, 402-408.

Leslie, A. M., Mallon, R., & DiCorcia, J. A. (2006). Transgressors, victims, and cry babies: Is basic moral judgment spared in autism? *Social Neuroscience*, *1*, 270-283.

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, *33*(2), 273-286.

Maccoby, E. E. (1992). The role of parents in the socialization of children: An historical overview. *Developmental psychology*, *28*, 1006.

Martin, G. B., & Clark, R. D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology*, *18*, 3.

Malle, B. F. (2004). How the mind explains behavior. *Folk Explanation, Meaning and Social Interaction.* Cambridge, MA*:* MIT Press.

Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: the consequences of disagreement. *Journal of Personality and Social Psychology*, *88*, 22.

Mikhail, J. (2000). Rawls' Linguistic Analogy: A Study of the 'Generative Grammar' Model of Moral Theory Described by John Rawls in 'A Theory of Justice.' (Phd Dissertation, Cornell University, 2000).

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*, 143-152.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge, UK: Cambridge University Press.

Moll, J., & Schulkin, J. (2009). Social attachment and aversion in human moral cognition. *Neuroscience & Biobehavioral Reviews*, *33*, 456-465.

Mullener, N., & Laird, J. D. (1971). Some developmental changes in the organization of self-evaluations. *Developmental Psychology*, *5*, 233.

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*, 221-236.

Nichols, S., & Folds-Bennett, T. (2003). Are children moral objectivists? Children's judgments about moral and response-dependent properties. *Cognition*, *90*, B23-B32.

Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, *52,* 114-121.

Nucci, L. P. (2001). *Education in the moral domain*. Cambirgde, UK: Cambridge University Press.

Nucci, L. P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child development*, 400-407.

Piaget, J. (1932/1965). *The moral judgment of the child. New York: Free Press*.

Prinz, J. J. (2006). *Gut reactions: A perceptual theory of emotion*. Oxford, UK: Oxford University Press.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57.

Railton, P. (1986). Moral realism. *The Philosophical Review*, *95*, 163-207.

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Belknap Press.

Rokeach, M. (1973). *The nature of human values* (Vol. 438). New York, NY: Free press.

Rozin, P. (1997). *Moralization*. Taylor & Frances/Routledge.

Rozin, P. (1999). The process of moralization. *Psychological Science*, *10*, 218-221.

Rusbult, C. E., Verette, J., Whitney, G. A., Slovik, L. F., & Lipkus, I. (1991). Accommodation processes in close relationships: Theory and preliminary empirical evidence. *Journal of Personality and Social Psychology*, *60*, 53.

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in the newborn. *Developmental Psychology*, *12*, 175.

Schein, E. H. (1985). *Organizational culture and leadership: A dynamic view*. San Francisco, CA: Jossey-Bass.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1-65.

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, *32*, 519-542.

Shweder, R. A., & Haidt, J. (1993). Commentary to feature review: The future of moral psychology: Truth, intuition, and the pluralist way.

Shweder, R., Much, N., Mahapatra, M., & Park, L. (1997). Divinity) and the" Big Three" Explanations of Suffering. *Morality and Health*, *119*.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy & Public Affairs*, 229-243.

Singer, P. (1981). *The expanding circle*. Oxford, UK: Clarendon Press.

Singer, P. (1995). *Animal liberation*. New York, NY: Random House.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: another contributor to attitude strength or something more? *Journal of personality and social psychology*, *88*, 895.

Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, *52,* 1333-1336.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. *Handbook of moral development*, 119-153.

Smith, P. B., & Schwartz, S. H. (1997). Values. In J. W. Berry, C. Kagitcibasi & M. H. Segall (Eds.), *Handbook of cross-cultural psychology, Vol. 3, 2nd Ed.* (pp. 77-119). Boston, MA: Allyn & Bacon.

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60,* 521-538.

Strayer, J., & Roberts, W. (2004). Empathy and Observed Anger and Aggression in Five‐Year‐Olds. *Social Development*, *13*, 1-13.

Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, *37*, 228-243.

Stone, J., & Cooper, J. (2003). The effect of self-attribute relevance on how self-esteem moderates attitude change in dissonance processes. *Journal of Experimental Social Psychology*, *39*, 508-515.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*, 345.

Thompson, R. A., Meyer, S., & McGinley, M. (2006). Understanding values in relationships: The development of conscience. *Handbook of moral development*, 267-297.

Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Psychology Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Turiel E., (1998). Moral development. In N. Eisenberg (Ed.) & W. Damon (Series Ed.), *Handbook of child psychology, Vol. 3: Social, emotional, and personality development* (5th ed., pp. 863-932). New York, NY: Wiley.

Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions, and vagaries. *The emergence of morality in young children*, 155-243.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., & Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in neural information processing systems* (pp. 1874-1882).

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological science*, *17*, 476-477.

de Waal, F. B. (1996). *Good natured* (No. 87). Harvard University Press.

de Waal, F. (1982. *Chimpanzee politics:* New York, NY: Harper & Row.

de Waal & Lanting, F. (1997). *Bonobo: The forgotten ape.* Berkeley, CA: University of California Press.

Wainryb, C., Shaw, L. A., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology*, *37* 373.

Wainryb, C., Shaw, L. A., & Maianu, C. (1998). Tolerance and intolerance: Children's and adolescents' judgments of dissenting beliefs, speech, persons, and conduct. *Child Development*, *69*, 1541-1555.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*(3), 271-294.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*, 780-784.

Whitbeck, L. B., & Gecas, V. (1988). Value attributions and value transmission between parents and children. *Journal of Marriage and the Family*, *50,* 829-840.

Williams, R. M., Jr. (1970). *American society: A sociological interpretation*, 3rd Ed. New York, NY: Knopf.

Wright, J. C., Cullum, J., & Schwab, N. (2008). The cognitive and affective dimensions of moral conviction: Implications for attitudinal and behavioral measures of interpersonal tolerance. *Personality and Social Psychology Bulletin*, *34,* 1461-1476.

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Comput Biol*, *4*, e1000254.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*, 8235-8240.

Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, *7*, 1-10.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*, 1912-1920.

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*, 202-214.

Zahn-Waxler, C., & Chapman, M. (1982). Immediate antecedents of caretakers' methods of discipline. *Child Psychiatry and Human Development*, *12*, 179-192.