

Morality by Tacit Agreement: A Contribution from the Economics of Emotions toward Moral Judgments

Abstract

Current research on morality is divided into rationalist and intuitionist theories. This study shows that when individuals make rational choices, they are inevitably guided by the moral foundation of intuitionism. Especially to pursue self-interest, individuals must agree with others in society. They must keep their opinions constant to agree with others. To maintain a constant opinion, the individual assigns an opinion that can improve the utility of the other person and place both of them in the same situation. The actions taken by an individual must be altruistic to enhance the utility of the other person. The individual's opinion is subsequently guided by the golden rule of morality, "Treat others as you would like to be treated," which facilitates moral foundations.

Keywords: Morality, Tacit Agreement, Moral Foundations, Golden Rule, Economics of Emotions

Introduction

In recent years, moral judgments have received considerable attention (Malle, 2021). However, the question arises, "Why do we have a moral sense?." Two contrasting streams of academic research answer this question (Garrigan et al., 2018). The first stream realizes the idea that moral judgments result from rational decisions. Exploring the logic of morality, many philosophers such as Hobbes (1651/2009), Locke (1689/1988), Kant (1785/1995), Rawls (1971), Gauthier (1987), and Scanlon (1998) have indicated that when humans make moral judgments, they behave as they had presumably bargained with others to reach an agreement about the distributing benefits and burdens of cooperation (Baumard et al., 2013). These contractarian philosophers argued that morality maximizes the mutual benefits of interactions. This trend, inspired by Piaget's (1932) view, was carried out by rationalists (or contractualists) such as Kohlberg (1981), who emphasizes that our moral judgments are driven by reasoning processes and are developed based on reflections and discussions over behaviors. Kohlberg highlighted the correlation between the individual

development of the logical and moral stages and proposed six stages of moral judgment divided into three levels: pre-conventional, conventional, and post-conventional.

The second stream involved Hume's (1751/1983) philosophy of morality. Hume believed that moral distinctions are not derived from reasoning processes but from institutions and that moral judgments are parallel to aesthetic judgments (Waldmann & Wiegmann, 2012). This stream has been continued by intuitionists (or nativists) such as Haidt (2011) who argued that the human mind innately possesses a moral sense. More specifically, the intuitionists' understanding of moral judgment has been developed as moral foundation theory (MFT), which arranges five foundations. Several studies have supported the existence of a direct intuitive link¹. For example, harmless taboo violations, such as eating a pet dog and a consensual incestuous relationship with birth control, are judged as morally wrong without providing reasons for their judgments. This kind of moral judgment without reason is called moral dumbfounding and provides evidence of the automatic and unconscious appearance of moral intuitions (Haidt & Hersh, 2001; Haidt et al., 1993)².

Parallel to these two contrasting streams of research is another research stream on morality based on economic rationality in the pursuit of self-interest (McClennen, 2010). Baumard et al. (2013) proposed a mutualistic approach based on partner choice that elucidates the motivation for moral behavior to maintain cooperative partnerships for mutual benefit. Although this is not explicitly tied to moral theory, Misyak et al. (2014) proposed virtual bargaining as a new theory of social decision-making. In virtual bargaining, individuals tend to coordinate their behavior based on what they would agree to do if they could bargain openly. Such tacit coordination based on virtual bargaining among individuals can be developed into unwritten rules or moral laws. Moreover, Levine et al. (2020) argued that when all individuals in a society suffer losses by performing a particular action, it becomes a shared moral law that no one should perform that action.

The economics of emotion (EoE), which attempts to model the human mind, is rapidly developing as AI advances; the EoE model is gaining importance in the study of the theory of mind. In addition, EoE has a high affinity with the study of moral judgments and proposes a valid hypothesis for "why" and "how" humans make moral judgments. The EoE model, from the perspectives of economic rationality and self-interest, derives a moral law that is analogous to intuitionists' moral foundation and argues that moral judgments necessarily lead to the golden rule, "Treat others as you would like to be treated." If the EoE shows that moral rules similar to moral foundations can be derived consequently from rational choice, it potentially integrates rationalist and intuitionist theories. Therefore, based on the EoE model, this study explains why humans are

¹ See Haidt (2001), Haidt & Kesebir (2010), and Haidt and Joseph (2011)

² Note that intuitionists do not ignore the role of reasoning, and it often provides post hoc rationalizations of moral intuitions that appear automatically and unconsciously.

endowed with a moral nature and introduces the mental mechanism of how moral judgments are made by socialized individuals. The study reveals that arguments concerning MFT are more than mere “hand-waving” but are inevitably rational choices available to all individuals (Haidt & Joseph, 2011; Suhler & Churchland, 2011).

The remainder of this paper is organized as follows. First, the discussion is based on the correspondence between situations and actions expressed by the form “in this case, this is what you should do.” Therefore, the correspondence between situations and actions is briefly introduced in the following section. Second, we reveal that by agreeing with each other’s opinions, an agreement is reached, and to achieve it, opinions must be kept constant. We explained how responsibility and rights accrue to individuals when trying to hold opinions constant to reach a consensus and why fulfilling responsibilities and protecting rights is significant. Finally, the Golden Rule is introduced, which states that when one tries not to change one’s opinion even if placed in the same situation as the other person, one tries to make the other person do what one wants rather than what one does not.

Motivation for agreement

EoE primarily assumes relationships or partnerships that last relatively long, such as friends, lovers, colleagues, and family members. It further considers that each individual, when placed in varied situations, responds variedly. In a situation, one satisfies a desire expressed in consciousness or constrain the necessary actions to fulfill a desire. The EoE model explains why and how individuals manifest morality in their consciousness and adjust their behaviors when their behaviors become an obstacle to their partners. In the present study, we briefly introduce the concept of situations in the EoE model and the behavior of individuals in response to these situations.

Situation and behavior

The discussion in the EoE model is based on the correspondence between the situation and optimal behavior. The situation is the problem to which the optimal behavior corresponds and responds. The situation in the EoE model comprises incentives, abilities, environment, and knowledge, each uniquely shaping an individual’s behavior. The EoE model assumes that the incentive causes the desire to appear in the individual’s consciousness, and, following the standard view of desire-satisfaction theory (McDaniel & Bradley, 2008; Schueler, 1995), the individual likely acts to satisfy the desire. However, an individual’s behavior is constrained by their abilities and the surrounding environment. Even if the actions satisfy the desire, an individual’s behavior will diversify with the differences in their abilities, the surrounding environment, and the knowledge possessed. Individuals generally fulfill their desires depending on their knowledge;

however, the knowledge available to individuals in a given situation is not always the same, even if they satisfy the same desire.

Perner and Wimmer's (1985) ice cream task stated below can elucidate how these elements shape individual behavior. Mary and John play in the park when an ice cream vendor visits it. Mary wants to buy ice cream. However, she does not have money and must go home to obtain her purse. While Mary is at home, the ice cream vendor moves from the park to church. John knows that the ice cream vendor is in the church, but he thinks Mary does not know about it. Therefore, if John tries to meet Mary, he will likely go to the park, not the church.

In the above example, when Mary perceives an ice cream van, the desire to eat ice cream triggers in Mary's consciousness. The situation for Mary begins when she sees the ice cream van or hears the sound of the vendor, both of which are incentives to trigger desire. However, concerning ability, if Mary has her purse with her when she tries to buy ice cream, she will not have to return to her house, and her ability will improve. Regarding the environment, if Mary has a purse but her mother forbids her from buying ice cream, Mary cannot take advantage of her ability, and her environment will deteriorate. Furthermore, Mary's decision about whether to go to the park or church to buy ice cream depends on whether Mary has knowledge that the ice cream vendor is at church. Mary's decision to go to the park or church alters according to the change in her knowledge and behavior toward buying ice cream.

In the EoE model, the situation in which an individual is placed comprises incentives that induce desires and abilities, environment, and knowledge that constrain behavior. When placed in a situation, the individual will act optimally to satisfy the desire generated by the incentive. However, apart from incentives, the action will be determined by abilities, environment, and knowledge that constitute the situation. The correspondence between the situation and optimal behavior will then be determined³. The correspondence in EoE elucidates an individual's opinion, which further defines the solution to the problem.

However, when individuals act optimally based on their free opinions, their actions may interfere with others' actions. When one's behavior is hindered by the other's behavior, one tries to reach an agreement on acceptable behavior for their situation.

Necessity of agreement in society

A society comprises multiple individuals who engage in various activities. Individuals who make the society inevitably tries to agree. In an interdependent situation, when one acts to satisfy

³ Until now, the three assumptions of rational choice theory have been derived: preference to incentive, constraints by ability and environment, and maximization of reward and minimization of cost (Opp, 2013).

one's desires, the action may prevent others from satisfying their desires⁴. The individual tries to avoid the behavioral condition hindered by other's behavior. Therefore, reaching an agreement with the other party on acceptable behavior for the situation becomes necessary. This agreement of opinion is the correspondence between the situation and the optimal behavior. In other words, agreeing with the opinion states, "This is what we should do in such a situation."

For example, Mary and John agreed to play together in the park. However, Mary went home to get her purse, preferring to eat ice cream rather than play with John. Mary's desire to eat ice cream can be fulfilled at the expense of John's desire to play with her in the park. John would then try to reach an agreement on the opinion that they should keep their promise of playing together in the park, even if they wanted to eat the ice cream. In another scenario, if Mary is late to meet John because she is engrossed in a TV program, she will make John wait. Mary's free fulfillment of her desire to watch the TV program prevents John from fulfilling his desire not to wait. John would then try to reach an agreement with the opinion that both of them should keep their promise not to be late for the meeting, even if they desire to watch the TV program. Thus, an agreement would be about what may and may not be done in a given situation.

Individuals also attempt to reach an agreement by satisfying each other's desires through cooperation. In several cases, the individual and their partner fulfill each other's roles to accomplish their desires. In this scenario, both encourage each other to fulfill their own roles. The agreement would then be on what roles to perform and in what situations.

For example, it is beneficial for Mary and John to play together in the park when neither can play alone. However, when one of them breaks their promise to play in the park by prioritizing reading a cartoon, both cannot play in the park. In this case, even if they wanted to read a cartoon, they would still try to reach an agreement to get together in the park.

Therefore, when individuals forming a society act freely to satisfy each other's desires, two types of desires can emerge simultaneously: the desire to avoid when individuals cannot satisfy their desires due to the actions of others and the desire to cooperate to satisfy their desires. The individuals inevitably attempt to reach an agreement on how they should act in a particular situation to satisfy their desires based on the mentioned two objectives. Shared agreement in society is defined as the correspondence between situations and actions.

Definition of opinion and agreement

This study introduces two purposes for agreement in society: to avoid being unable to satisfy one's own desires or cooperate to satisfy one's desires with others. Agreement in the EoE model

⁴ Literature concerning economic rationality in both independent and interdependent situations is comprehensively discussed by Thielmann et al. (2020).

is established when there is consensus among individuals. Opinions in the EoE model are defined as desirable actions in a specific situation or what actions should be taken under what circumstances. Therefore, this section introduces a definition of agreement in the EoE model based on the definition of opinions.

Unanimity of opinion and agreement

The correspondence between the situation and action is called an opinion when an individual decides what action should be taken in response to a particular situation. For example, when one creates an opinion about a situation, one contemplates what to do if the other person is in the same situation⁵. The individual decides what action should be taken in response to the other person's situation, and an opinion is created by the correspondence between the situation and action. In other words, the individual will virtually place themselves in the same situation as the other person.

By perceiving the same incentives as the other person, the other person's desires will appear in the individual's consciousness, and the individual can share the same goal of "What do you want to do?" However, when confronted with the same abilities, environment, and knowledge as the other party, the individual can be constrained to the same action and share the same ways and means with the other party regarding "What should you do?."

For example, when Mary tries to satisfy her desire for ice cream, John can make sense of Mary's behavior of returning to her house, getting her purse, and buying ice cream. The desire to eat ice cream at that moment allows him to understand Mary's goal of "What does she want to do?." In addition, by Mary's action of returning home to get her purse, John can understand her method of "What should she do?."

Therefore, when deciding on an opinion, an individual extensively increases rewards and decreases costs related to the situation when the individual's actions do not affect the other party's actions. Opinions can then be expressed as "in such a case, one should do this." In other words, the expression "in such a case" states what is to be done in a specific situation, and the expression "one should do this" expresses the action to be taken in response to that specific situation. However, when creating opinions on the same situation, the individual can compare whether both partner's opinions are the same. We call this same opinion a unanimous opinion agreement⁶. Both parties must meet minimum conditions to reach an agreement.

⁵ This is consistent with simulation theory (ST). In ST, one thinks about what one should do by putting oneself in the other person's situation (Goldman, 2006).

⁶ When the individuals who make up a society make an agreement, the correspondence between situations and actions is shared, and a norm is established in the society (Opp, 2013).

Uniqueness of opinion

To agree on an opinion, the individual and their partner must not change the correspondence between the situation and the action. When either of their opinions changes, agreement cannot be accomplished. For example, with a proposed opinion that action A should be taken in response to a particular situation, both the individual and their partner will reach an agreement. However, when actually placed in that situation, the partner may consider action B. If the partner changes their opinion, reaching an agreement is impossible as an agreement is no longer available between both parties. Therefore, the minimum condition to reach an agreement is that neither party must change their opinions. Keeping opinions unchanged and constant is called the uniqueness of opinion.

For example, when Mary and John agree that they should keep their promise to play in the park, even if they want to have ice cream or watch a TV program, both will be in agreement. To satisfy the uniqueness of opinion, they must not prioritize having ice cream or watching TV programs, even if they desire. Mary and John also agree that they should inform the other person that the ice cream vendor has moved to church. To satisfy the uniqueness of opinion, they must make an effort to teach the other person, even if they pay the price of teaching.

Every individual has a motive to try to satisfy the uniqueness of opinion. Without satisfying the uniqueness of opinion, one cannot agree with others to reach a consensus. The individual can be excluded from the partnership, community, society, and so on when judged incapable of agreement, which is a worse outcome than following the agreement (Baumard et al., 2013). When individuals seek to improve their utility, they inevitably agree with others and act according to the agreement. Rationality in EoE thus becomes visible. While this rationality is consistent with that of the mutualistic approach, it differs in that the EoE emphasizes maintaining uniqueness in one's opinion to reach an agreement with others. Therefore, when agreeing with others, the individual must not change their opinion once stated, "This is what we should do in such and such a case." Moreover, while avoiding changing one's opinion, a motivation to act following one's own opinion transpires, even if that opinion has negative consequences.

Individuals who change their opinions about "what should be done in such a situation" will have their partnerships dissolved or excluded from the community. They are evaluated as incapable of reaching an agreement by satisfying the uniqueness of their opinions. For example, criticisms such as "what you said then and what you say now are different," "what you say and what you do are different," and "you are hard on others but soft on yourself" indicate that the opinions do not meet uniqueness.

In addition, the other persons must not want the same thing to be done as they did to the one, and their opinions vary. In this case, the uniqueness of the opinion is not satisfied, which is the minimum requirement for agreement. Therefore, the other person will try to do things that would

not bother them if the one did the same to them. An individual tries to adjust their opinion because they do not want to change it even if they are later placed in the same situation as the other party.

Commutability of opinion

To satisfy the uniqueness of opinion for oneself and one's partner to reach an agreement, one necessarily fulfill the following conditions. First, one tries to avoid changing one's opinion even if placed in the same situation as the other person. For avoidance, one proposes an opinion to the other person in advance that satisfies the one's desires as much as possible. In other words, the individual satisfies their desires similarly to the other person in the same situation based on the agreement. Therefore, the individual will not have to change their opinion, fulfilling the uniqueness of opinion.

Second, if one does not allow the other party to act to satisfy the other party's desire, when one is placed in the same situation as the other party, one will not be able to act to satisfy one's desire in the same way as the other party. Therefore, to reach an agreement with the other party, the individual inevitably tries to make the other party act the way one wants in that situation. An opinion that can satisfy the individual's desire in advance to place the individual in the same situation as the other person is called commutativity of opinion⁷.

For example, if John wants Mary to prioritize playing in the park over eating ice cream and watching a TV program, John must prioritize playing in the park rather than eating ice cream and watching the TV program. Conversely, if John wants to prioritize eating ice cream or watching TV, he must let Mary prioritize the same. Only then will John satisfy commutativity (uniqueness). Moreover, when John informs Mary that the ice cream shop has moved, he will be able to ask Mary to do the same. If John does not want to bear the burden of informing Mary, he cannot accuse Mary of doing the same thing.

In this case, agreement in society will not necessarily be explicitly confirmed but assumed implicitly. For example, when Mary and John avoid going home and being on time for a meeting, they will not confirm each other's opinion on avoiding going home and being on time for the meeting. When it is assumed that they should maintain the meeting time, the implicit agreement becomes tacit. Moreover, when John informs Mary that the ice cream shop has moved, Mary may willingly provide John with the same useful information. When Mary and John provide each other with useful information without explicitly exchanging opinions, they are in tacit agreement.

⁷ This commutativity must be satisfied in opinion, especially when there is a possibility that one will be placed in the same situation as the other person in the future. Thus, the commutativity of opinion is especially important in communities that share the same identity and require solidarity (Metz, 2021).

Many societal agreements are tacit. For example, many customs, common sense, etiquette, and manners shared within a community are tacit agreements. Tacit agreements are defined as correspondences between situations and actions in the form of “this is what we should do in such and such a case.” In the behavioral expression, “this is what we should do,” desirable and undesirable behaviors are distinguished based on the tacit agreement. In addition, in teaching customs, common sense, etiquette, manners, etc., the form “this is what we should do in such and such a situation” is shared. When following the opinion of “this is what we should do in such and such a case” is beneficial for both parties, it becomes rational for all individuals who constitute a society to follow the customs, common sense, etiquette, and manners. For example, when we are punctual during a meeting or do not make noise while eating, we do not have to wait or feel uncomfortable while meeting or eating. Contrastingly, an explicit agreement is a clearly stated law or social institution. For example, when we do not harm others or violate their rights to their property, we can ensure our physical safety and protect their property. In addition, when we use public facilities with care and walk on the right side of the road, we can use the facilities at a lower cost for a long and will not have to pay unnecessary attention when walking.

Explicit and tacit agreements can also be made regarding various roles and responsibilities. For example, in a company or sports team, different roles are assigned to each team member. As long as all teammates, including oneself, benefit by fulfilling one’s role, one will receive an incentive. To ask others to fulfill their roles to achieve results as a team, fulfilling one’s own role according to the uniqueness of opinions is necessary.

Responsibilities and rights by tacit agreement

Individuals attempt to act freely to satisfy their desires. When an agreement is not shared in society, an individual may less satisfy their desires than when the agreement is made. For example, without a shared agreement in society, people will commit murder, robbery, theft, fraud, etc., to satisfy their desires. Contrastingly, when people follow the agreement shared in society, they will not be victims of other individuals. Thus, for an individual, agreeing with others, even if this means restricting one’s behavior, is often a better option. Responsibility and rights are created in society only when one tries to get the other party to follow an agreement in exchange for one’s compliance.

Roles and positions by agreement

When an individual and the other party reach an agreement on the action to be taken in response to a situation, both try to act according to the agreement and get the other party to do the same. When an agreement determines the action to be taken, an individual’s role is determined. An individual’s position is also determined when getting the other party to act according to the

agreement. Role determination ensures that one will prioritize fulfilling that role according to an agreement rather than freely fulfilling desires. Moreover, while fulfilling one's role according to the agreement, the desire to fulfill the role will be similar to the one that appears due to the incentive, and that desire determines one's responsibility. When one's position is determined, one tries to get the other party to prioritize acting according to the agreement rather than allowing the other party to fulfill one's desires.

In other words, when an individual asserts their position, they get the other party to prioritize fulfilling their role according to the agreement, even if this prevents the other party from freely fulfilling their desires. While fulfilling a role, the desire to have the other party fulfill one's role appears in one's consciousness. This desire to have the other party fulfill a role is called a right. Individual responsibilities and rights are manifested based on agreements. Responsibility and rights are justified because when individuals act following their mutual agreement, they will not prevent and will better fulfill each other's desires.

However, a conflict can arise between the desire to fulfill a role and the desire expressed by incentives where one must sacrifice their desire to favor the other. In other words, when an individual prioritizes satisfying their desires, the worst outcome may be to break an agreement with the other party, making it impossible to reach further agreement. In contrast, when an individual prioritizes keeping the agreement made with the other party, they restrict their behavior, making it impossible to fulfill their desires freely. When a conflict arises between fulfilling the desires expressed by incentives and one's role according to an agreement, one's partner must coordinate by discussing how much priority should be given to which situation. When one's partner agrees on acceptable behavior for the situation, both parties can reach an agreement.

Individuals placed in a situation freely satisfy their desires when the desire is expressed by the incentive. However, when the other person is placed in the same situation, the individual must also allow the other person to fulfill their desires freely. In addition, the limit of fulfillment of the desire manifested by an incentive in a situation is based on the agreement. Therefore, if an individual limits their fulfillment of desires based on the agreement, the other person can also be stopped from satisfying their desires based on the agreement.

For example, when given a position in a company, an individual, along with all employees, agrees to prioritize fulfilling their job responsibilities even if they have other important personal works or conflicts of interest. When they fulfill their assigned roles, they can map their agreed-upon actions to the situation. By doing so, if the other person is placed in the same situation as the individual, the person may be asked to take actions the individual has agreed upon. In this case, both are performing moral actions by trying to fulfill one's responsibilities and protecting the rights of the other person.

Moral foundations by agreement

When individuals try to satisfy their desires in society, they inevitably try to reach an agreement with other individuals. While reaching an agreement, they try to satisfy the uniqueness (commutativity) in their opinions. The uniqueness of opinion is satisfied when an individual makes the other person do something that the individual wants to do. In addition, the individual must do to the other person what they want the other person to do to them. When one does something that one wants must be done by the other person when both are placed individually in the same situation, that action is an attempt to fulfill one's responsibility. In this case, one's action will manifest itself as an altruistic behavior toward the other person. Further, one's actions toward the other person will be an attempt to protect one's rights, which requires altruistic behavior from the other person. Thus, altruistic behavior manifests when an individual tries to reach an agreement with another party by satisfying the uniqueness of opinion. Therefore, by satisfying the uniqueness of opinion, an individual necessarily follows the golden rule, "Treat others as you would like to be treated."

Moral foundations can be derived from tacit agreements in the EoE. In essence, moral foundations appear when one tries to satisfy the uniqueness of opinion to reach an agreement with the other party. Specifically, regarding care/harm, if one wants to be cared for or not be harmed, based on the uniqueness (commutativity) of opinions, one must care for or not harm the other party. Regarding fairness/cheating, when both parties reach an agreement, both must take the same action in the same situation and satisfy the uniqueness of their opinions. Furthermore, for loyalty/authority, as long as both parties benefit, there will be an incentive to fulfill the responsibilities according to the roles assigned to each party, and loyalty and authority will then be morally justified. Finally, regarding sanctity/degradation, if one desires to treat one's partner in a certain way, one must be prepared to be treated in the same way by the partner, or else one cannot satisfy the uniqueness of opinion.

This conception of morality, based on tacit agreements, explains the process of moral development. Individuals are placed in various situations while growing up, which helps them to form opinions about what they should do in different situations. Therefore, in the process of personal growth, individuals can form their opinions about various situations and develop their own way of thinking. Once there is an agreement among individuals, any kind of agreement can be reached. When morality is expressed based on agreement, it can differ among societies while agreeing on different opinions. Thus, morality based on tacit agreements can differ both culturally and regionally.

Conclusion

This study introduced how individuals who make up society can derive moral principles, such

as moral foundations when they try to reach an agreement. This study paves the way for in-depth research on how an individual's altruistic behavior can enhance the utility of the other person for a more agreeable society, guided by the golden rule of morality, "Treat others as you would like to be treated," which derives the individual's moral foundation. Also, the study may help researchers and educationists focus on how morality based on tacit agreements that differ both culturally and regionally influence social well-being as a whole.

In addition, the model introduced in this study has potentially answered the "why" and "how" questions of morality (Baumard et al., 2013). The "how" question is about the mental mechanisms of how people make moral judgments and interactions. In response to this question, the EoE model provides a criterion that one should do to the other what one wants the other to do to them, based on the commutativity of opinions. The "why" question is about the reason why humans have morality. Responding to this question, the EoE model suggests that opinions will change unless their uniqueness is satisfied, making it impossible to reach an agreement with others or cooperate to improve utility. Therefore, human moral law is based on rational thinking. When an individual feels anxious about not being able to agree with others or a sense of discomfort arises when they do not satisfy their rationality, guilt and conscience appear in consciousness. We believe that moral law is innate in human beings by intuitionists because of the unawareness of rationality, as presented in this study, and it turns out to be incorrect to consider that moral law is intuitive when rationality is clearly understood.

References

- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36(1), 59-78.
- Garrigan, B., Adlam, A. L. R., & Langdon, P. E. (2018). Moral decision-making and moral development: Toward an integrative framework. *Developmental Review*, 49, 80-100.
- Gauthier, D. (1987). *Morals by agreement*. In: Clarendon Press/Oxford University Press.
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.
- Haidt, J., & Hersh, M. A. (2001). Sexual Morality: The Cultures and Emotions of Conservatives and Liberal. *Journal of Applied Social Psychology*, 31(1), 191-221.
- Haidt, J., & Joseph, C. (2011). How moral foundations theory succeeded in building on sand: A response to Suhler and Churchland. *Journal of Cognitive Neuroscience*, 23, 2117-2122.
- Haidt, J., Roller, 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(4), 613-628.

- Hobbes, T. (1651/2009). *Leviathan*. Oxford University Press.
- Hume, D. (1751/1983). *An Enquiry Concerning the Principles of Morals*. Hackett Pub Co Inc.
- Kant, I. (1785/1995). *Ethical Philosophy: Grounding for the Metaphysics of Morals & Metaphysical Principles of Virtue*. Hackett Pub Co Inc.
- Kohlberg, L. (1981). The Philosophy of Moral Development Moral Stages and the Idea of Justice In. San Francisco, CA: Harper & Row.
- Levine, S., Kleiman-Weiner, M., Schulz, L., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences*, *117*(42), 26158-26169.
- Locke, J. (1689/1988). *Locke: Two Treatises of Government Student edition*. Cambridge University Press.
- Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology* *72*, 293-318.
- McClennen, E. F. (2010). Rational Choice and Moral Theory. *Ethical Theory and Moral Practice*, *13*(5), 521-540.
- McDaniel, K., & Bradley, B. (2008). Desires. *Mind*, *117*(466), 267-302.
- Misyak, J. B., Chater, & Nick. (2014). Virtual bargaining: a theory of social decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130487.
- Opp, K.-D. (2013). Norms and rationality. Is moral behavior a form of rational action? *Theory and Decision*, *74*, 383-409.
- Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that...": Attribution of Second-Order Beliefs By 5- to 10-Year-Old Children. *Journal of Experimental Child Psychology*, *39*(3), 437-471.
- Piaget, J. (1932). *The moral judgment of the child*. The Free Press.
- Rawls, J. (1971). *A theory of justice*. Belknap Press: An Imprint of Harvard University Press.
- Scanlon, T. M. (1998). *What we owe to each other*. Belknap Press of Harvard University Press.
- Schueler, G. F. (1995). *Desire: Its role in practical reason and the explanation of action*. MIT Press.
- Suhler, C. L., & Churchland, P. (2011). Can Innate, modular "foundations" explain morality? Challenges for Haidt's Moral Foundations Theory. *Journal of Cognitive Neuroscience*, *23*(9), 2103-2116.
- Waldmann, M. R. N., Jonas, & Wiegmann, A. (2012). Moral Judgment 19. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 364-389). Oxford University Press.