Universitat
de les Illes Balears

DOCTORAL THESIS
2019

THE SECOND-PERSONAL DIMENSION OF OUR MORAL PSYCHOLOGY

Carme Isern Mas

Universitat
de les Illes Balears

EvoCog IFISC

DOCTORAL THESIS
2019

Doctoral Program of Cognition and Human Evolution

THE SECOND-PERSONAL DIMENSION OF OUR MORAL PSYCHOLOGY

Carme Isern Mas

Thesis Supervisor: Antoni Gomila
Thesis Tutor: Antoni Gomila

Doctor by the Universitat de les Illes Balears

A Isabel Segura,

Acknowledgments

Les persones que em coneixen saben la il·lusió que em fa escriure aquests agraïments. Els he estat pensant durant quatre anys, que han passat tan ràpid que espanta. I són tantes les persones per qui sento gratitud, que espero que la meva memòria no em jugui una mala passada ara que els toca a elles rebre tot el mèrit que es mereixen. Comença aquí doncs una de les meves llistes més esperada: la de les càlides i merescudes gràcies a totes aquestes persones.

Un dels pilars importantíssims d'aquests quatre anys han estat (i això no sorprendrà a ningú): els meus increïbles companys de feina d'Evocog i de l'equip veí de Neurocog. Bé, "companys" no; amics. Amics de feina. La nostra ha estat la sèrie més bonica en la qual mai he actuat. Hem tingut episodis amb una mica de tot: molt honorables representants, *challenge*, "doctorakis y detectives", noces, Cala Canta, karaokes, Pina, pàdel, Meninas, San Fermín amb "*Darwín* hasta el fin"... I podria seguir durant pàgines i pàgines, però per respecte a nosaltres mateixos ho deixaré aquí. Ho he dit des del primer dia i ho he reafirmat en aquests quatre anys: molt bé van les coses si després de passar més de 40 hores a la setmana amb unes persones encara les vols seguir veient durant les hores restants. Ho he agraït cada dia, sobretot quan he parlat de vosaltres, però ho vull deixar per escrit: Albert, Alicia, Ana, Conchi, Cris, Daniela, Emilio, Erick, Guido, Hugo, Irune, Pame, Quique; gràcies.

Gràcies a la resta de professors-investigadors d'Evocog i afins (Camilo Cela-Conde, Enric Munar, Fabrice Parmentier, Jaume Rosselló, Jordi Pich, Juan Tomás Escudero i Marcos Nadal) pels consells, les ensenyances i la proximitat. Gràcies especials a Toni; per recolzar-me amb la beca, per alegrar-te de què l'obtinguéssim i per ajudar tant i en tantes coses. Gràcies per posar-me límits quan ha calgut, per no posar-me'n quan jo els volia però no els necessitava, per deixar-me arribar a les conclusions a jo soleta, per la paciència, per la comprensió i per tota la feina. Gràcies.

En realitat aquesta tesi ha estat possible per la fantàstica feina de moltes persones que m'han educat i ensenyat, des de molt abans dels inicis de la tesi. Al final, tot el que sé (i sóc?) és per elles. Gràcies a les mestres de l'escola d' Es Molinar (Bel Frau, Carme Rosillo, Carme Salvà...) i gràcies al professorat de l'IES Josep Maria Llompart (Andreu Estarellas, Antònia Amengual, Antònia Ordinas, Catalina Cerdà, Catalina Manera, David Ginard, Francesc Artigues, Francina Flexas, Joan Obrador, Manuela Fuentes, Margalida Gelabert, Mateu González, Victòria López, Xavier Valentí, Xisca Bosch i tants altres). De tots vosaltres he après, gràcies. Gràcies al professorat de filosofia de la Universitat de les Illes Balears i de la Universitat de Barcelona, que van inspirar inconscientment aquesta tesi; i gràcies al professorat de psicologia de la UIB, que inconscientment l'han vista créixer. Gràcies especials a Bernat Riutort, Lucrecia Burges, Tomeu Sales i Joaquin Valdivielso per la confiança i el suport a l'hora d'impartir docència a les vostres assignatures. Gràcies també als meus professors d'anglès, Chris, Mikel, Brenda i Bernat, per una ajuda essencial.

Gràcies als alumnes del grau de filosofia, als alumnes del programa MentoriMENT i als alumnes del festival Inca Pensa (gràcies a l'organització del programa i del festival per confiar en mi, especialment a Maria Antònia Perelló i Pilar González per pensar en mi). Si algun alumne ha gaudit d'alguna de les classes una dècima part del què he gaudit jo fent-les; l'esforç ha valgut la pena.

A més d'escriure tesi i donar classe, en aquests anys he tingut el plaer de viatjar molt. En la meva primera estada: University of Antwerp. Gràcies a Jens, Laura, Lotte, Marco, Michael, Martha i Sarah per la càlida acollida; i gràcies a Jasper pels "breaks", per La La Land i per Frasier. Moltíssimes gràcies a la persona que ho va fer possible i que tant em va ajudar en aquella estada: Katrien Schaubroeck, *honestly, thank you.* Gràcies també a les meves boniques companyes de pis (Andrea, Anja, Sarah i Tatiana).

Per a la segona estada: Yale University. Gràcies a Stephen Darwall; *for accepting my application, for your time, for your teachings, and for your patience.* Gràcies a Joshua Knobe, Shelly Kagan, Paul Bloom i Laurie Santos pel seu temps.

Gràcies a totes les persones que em van escoltar i guiar (Alain, Austen, Jake, Jennifer, Jo, Joanna, Juan, Kate, Lea, Marissa, Mike, Moya, Will...) i un gràcies especial a Brian, Jan i Mario pels cafès, els passejos i els consells. Gràcies a la imprescindible comunitat internacional amb la qual vaig compartir l'estada (Dries, Huub, James, Jelscha, Min i Nic) i un gràcies càlid i especial a la meva "germana" i persona de contacte per a les emergències, Laura. Gràcies també als meus companys de pis per fer-ho tot més fàcil, especialment a Alyka, Bobby, Guillaume, Sounak i Spencer. Gràcies també a dues persones que, tot i ser de fora de Yale, han estat part essencial de l'estada: Tommy, gràcies per reaparèixer (o per no haver desaparegut mai); i Will, gràcies per aparèixer.

Gràcies també a les persones que he conegut per congressos, workshops i escoles (d'estiu, de tardor, d'hivern i de primavera). En especial, gràcies a Fernando Broncano, Guillaume Dezecache, Hugo Viciana, Mark Alfano, Maureen Sie, Richard Moore i Sven Nyholm pels consells i la dedicació extra.

Més enllà del món acadèmic, però encara dins de la Universitat; gràcies als fantàstics cambrers del Guillem Cifre per fer-me riure; gràcies a Raquel per netejar-nos el despatx molt, molt d'hora; gràcies als conserges per ser sempre el primer i últim somriure del dia; gràcies als de seguretat que ens obren la porta en dies "no lectius" i gràcies a Cati Miralles i a Miquel Juan per l'ajuda en les tasques administratives.

Més enllà de la universitat, un altre dels pilars d'aquests quatre anys ha estat l'esport. Gràcies als entrenadors de Campus Esport (Ivan, Lidia, Sandra i Paloma) i un gràcies especial a Mikie per passar d'entrenador a amic. Gràcies a la gent que fa aquell gimnàs acollidor (Carmes, Dani, Elena, Fede, Kika, Mercè, Neus, Pedro...). Gràcies a Sandro, del gimnàs Het Ein Eiland (a Antwerp) i gràcies a Dan, del Crossfit Science Park (a New Haven). Gràcies als entrenadors del Pure Fitness Gym: Erik, Israel i Mireia, gràcies per la paciència i els consells. Gràcies als apassionats Jesús i Neus, del recent fundat equip OCR Warriors i gràcies als membres de l'equip per convertir-se en això, en un equip. Gràcies, África, Alba, Andrea, Carlota, Dani, Gris, Isa, Jose, Juanpe, Óscar, Pablo, Pilar, Sara i

Yennifer. I un gràcies immens i merescudíssim a les dues persones que em fan la vida més fàcil (a dins i fora del gimnàs): Pep i Mónica, per tantes coses, gràcies.

L'altre gran pilar al marge de la universitat han estat els infants; els dels esplais de l'Encarnació, Xiroia i Sant Medir, i els del City Plaza; sou la millor desconnexió i injecció d'energia. Gràcies als voluntaris, tant de l'esplai com de City Plaza, per la dedicació i gràcies als cuidadors dels infants per tantíssima estima.

Suport incondicional per aquests anys han estat els meus companys de pis. Gràcies en primer lloc a Patrick per atrevir-se a marxar amb mi i per haver estat un suport essencial en aquests quatre anys. Gràcies també a Carme i a Sarah, per unir-se a l'aventura amb tanta tendresa. Gràcies als habitants de Jumangi; Alva, Nuba, Petit i Thor. Gràcies a Joan Mateu i Laura per acollir-me en tornar dels Estats Units. Gràcies per la comprensió, per la companyia i pel suport.

Gràcies, per suposat, a les amistats. Gràcies a les amistats que ja hi eren des de molt abans de la tesi. Gràcies a "ses des coro" (Aina, Concha, Mar, Maria, Neus i Núria); per seguir sent i estant, per estimar i per sempre sumar. Gràcies a "ses d'Ibitza" (Aina, Alejandra, Carme i Lluc), pels sopars amb acta acordada, per la comprensió i pel suport. Gràcies, Bàrbara, per reaparèixer amb tanta força; gràcies, Mary, per entendre-ho *tot*; gràcies, Lluís, per no deixar de trucar; i gràcies, Mercè, per l'escolta, la confiança i els bons consells. Gràcies també a les amistats que des de lluny segueixo sentit a prop. Gràcies Albert, Arnau, Clara, Mickaël, Marcan, Miguel, Pablo; gràcies per les trucades, pels *skypes*, per les visites, pels viatges i per tant més.

Gràcies a les amistats que han aparegut durant la tesi. Gràcies als de "la hamburguesa del mes del Brox" (Alex, Andrés, Toni, Marta, Mikie, Mónica i Pep) per donar-me una bonica excusa per sopar entre rialles un cop al mes; gràcies als meus companys del grup de tardes de segon de grau psicologia del curs 2016-17 per fer-me part del grup; i un gràcies d'última incorporació però igualment essencial, als meus companys del màster del

professorat per fer els horabaixes més fàcils. Gràcies, especialment a Adrià, Miquel i Zoe per l'estima, l'afinitat i tantíssim de suport (no només en la tesi).

Gràcies a la família, a la propera i a la llunyana. Gràcies a la família "llunyana" (som massa cosins, cosines, oncles i ties per a anomenar-vos tots!) per l'interès en i el respecte a la meva feina. Gràcies a les persones que més he enyorat durant les estades; pare, mare i Laura. Gràcies per estar pendents de mi, per esperar-me, per acompanyar-me, per respectar les meves decisions i per escoltar-me (i animar-me) sempre que ho he necessitat. Gràcies per sempre, sempre, sempre ser-hi. Gràcies imprescindibles també a l'única persona que no ha vist la tesi acabar, però que *literalment* la va somniar; gràcies, iaia, per ser inspiració, exemple, motivació, llum, lluita, generositat i vida. Tot el què hi ha aquí et pertoca: perquè de tu ja ho vaig aprendre tot. Gràcies.

Tot i l'incommensurable suport emocional que he rebut, res no hagués estat possible sense les adequades condicions materials. Gràcies per això al Ministeri d'Educació, Cultura i Esport d'Espanya (ara Ministeri de Ciència, Innovació i Universitats) per concedir-me la beca FPU (FPU14/01186) i per finançar les meves estades a les universitats d'Antwerp (EST16/00260) i Yale (EST17/00141). I gràcies al Ministeri d'Economia i Competitivitat (ara Ministeri de Ciència, Innovació i Universitats) per concedir els projectes dels quals he format part en aquests anys (FFI2013-44007-P i FFI2017-86351-R).

Scientific papers related to the contents of this dissertation

Isern-Mas, C., & Gomila, A. (2019). Why does empathy matter for morality?. *Análisis Filosófico*

Isern-Mas, C., & Gomila, A. (*in press*). Segunda persona y moralidad. En *Perspectivas en Filosofía de la Subjetividad.* Departamento de Filosofía de la Universidad de la Concepción (Chile)

Isern-Mas, C., & Gomila, A. (2019). Making sense of emotional contagion. *HUMANA.MENTE Journal of Philosophical Studies, 12*(35), 71-100. https://www.humanamente.eu/index.php/HM/article/view/209

Isern-Mas, C., & Gomila, A. (2018). Externalization is common to all value judgments, and norms are motivating because of their intersubjective grounding. *Behavioral and Brain Sciences, 41*, E104. http://dx.doi.org/10.1017/S0140525X18000092

Isern-Mas, C. & Gomila, A. (2015). An Attempt to Understand the Relation Between Emotional Contagion and Motor Mimicry. *Proceedings of the 8 SLMFCE*

Isern-Mas, C. & Barone, P. (2015). La importancia de la mirada social en la comprensión de la mente ajena. *X Boletín de Estudios de Filosofía y Cultura Manuel Mindán.*

Conference presentations related to the contents of this dissertation

Isern-Mas, C. & Gomila, A. (2019). A second-person approach to the evolution of morality. Talk presented at the *Summer School Ethics, Empathy and Errors*, Pühajärve (Estonia)

Isern-Mas, C. & Schaubroeck, K. (2017). Why Does Morality Bind Us? The Role of Care in Moral Motivation. Talk presented at the *OZSW Conference*, Doorn (The Netherlands)

Isern-Mas, C. & Gomila, A. (2017). Love can ground Morality. Talk presented at the *Second Antwerp Summer School in Philosophy and Society: The Right to be Loved*, Antwerp (Belgium)

Isern-Mas, C. (2017).The Scope of the Moral Community. Discussant at the *12th Nomos Meeting: Symposium on Stephen Darwall's Second Personal Approach to Morality*, Palma (Spain)

Isern-Mas, C. & Gomila, A. (2016). Looking Into the Sense of Justice. Talk presented at the *WS on Experimental Philosophy: Methods and New Directions*, Berlin (Germany)

Isern-Mas, C. & Gomila, A (2016). Why Does Empathy Matter for Morality?. Talk presented at the *ESPP Conference*, Saint Andrews (Scotland).

Isern-Mas, C. & Gomila, A. (2016). Discussing Whether Empathy is Necessary for Morality. Talk presented at the *CERE Conference*, Leiden (The Netherlands).

Isern-Mas, C. & Gomila, A. (2016). Discussing Whether Empathy is Necessary for Morality. Talk presented at the *EPSSE Conference*, Athens (Greece).

Isern-Mas, C. & Gomila, A. (2015). Making Sense of Emotional Contagion. Talk presented at the *ESPP Conference*, Tartu (Estonia).

Isern-Mas, C. & Gomila, A. (2015). An Attempt to Understand the Relation Between Emotional Contagion and Motor Mimicry. Talk presented at the *VIII Conference of the Spanish Society for Logic, Methodology and Philosophy of Science*, Barcelona (Spain).

Isern-Mas, C. & Gomila, A. (2015). Un modelo para entender la relación entre el contagio emocional y la imitación motora. Talk presented at the *IX Psychology Students National Conference,* Valencia (Spain)

UIB Universitat de les Illes Balears

Dr. Antoni Gomila of the University of the Balearic Islands

I DECLARE:

That the thesis titled “The second-personal dimension of our moral psychology”, presented by Carme Isern Mas to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, I hereby sign this document.

Signature

A Gomila

Dr. Antoni Gomila Benejam

Palma de Mallorca, 10 October, 2019

Resumen

En nuestro día a día, los humanos realizamos muchas conductas altruistas: ayudamos a los amigos y a las amigas que nos necesitan, donamos sangre, hacemos voluntariados con los niños y las niñas de nuestro barrio, y damos dinero para buenas causas. Realizamos todas estas conductas con la motivación de ayudar a los demás o como resultado de nuestra consciencia sobre qué es lo correcto. De alguna forma, sentimos que no podemos sino actuar según nuestros juicios morales, pues nos sentimos especialmente comprometidos a actuar según ellos. Este tipo de motivación que emerge de nuestros juicios morales es lo que en filosofía y psicología moral se ha denominado "motivación moral". Este es el objeto de estudio de esta tesis.

El capítulo 1 se centra en el debate sobre si la motivación moral existe. Presentamos los argumentos y la evidencia empírica principales en contra de la existencia de la motivación moral y presentamos también los argumentos y la evidencia empírica a favor. Para defender muestra posición nos centramos en el sentimiento de culpa como indicador del papel de la moralidad para motivar y guiar nuestras acciones.

El capítulo 2 describe nuestra psicología moral en términos de nuestra naturaleza social, relacional. Proponemos una interpretación naturalista del trabajo analítico de Stephen Darwall en el libro *The Second Person Standpoint* (2006). Esta interpretación nos permite tratar algunos de los temas principales en psicología moral, como la motivación moral, las obligaciones especiales, las emociones morales, los límites de la comunidad moral o la emergencia de las normas de grupo.

El capítulo 3 responde a la pregunta sobre cómo funciona la motivación moral en el contexto de la amistad. En contra de Kant y Darwall, defendemos que la motivación moral y las obligaciones morales pueden surgir de este tipo de relación interpersonal. Específicamente, defendemos que la amistad implica una motivación especial, a saber, motivación para la parcialidad; que la amistad *requiere de* ese tipo de motivación; y que la amistad implica obligaciones especiales e implícitas.

El capítulo 4 se centra en la pregunta sobre evolución de la moralidad, concretada con respecto al origen de la motivación moral. Proponemos una explicación evolutiva unitaria de la moralidad que pueda explicar lo que consideramos que son las dos principales clases de motivaciones para actuar moralmente: motivación por parcialidad y motivación por imparcialidad. Retomamos la perspectiva de segunda persona naturalista de la moralidad introducida en el segundo capítulo y defendemos que esta versión naturalista puede explicar la emergencia evolutiva de los dos tipos de motivaciones morales. Argumentamos que nuestra propuesta puede evitar las limitaciones de la explicación evolutiva de la moralidad de Michael Tomasello (2016).

Los capítulos 5 y 6 exploran las implicaciones de nuestra fundamentación de la motivación moral en la interacción intersubjetiva en otros debates. Estos capítulos han sido publicados como artículos independientes. El capítulo 5 se presenta como una respuesta a la posición de Jesse Prinz en contra de la empatía. Defendemos que si la moralidad y la empatía se entienden debidamente, el papel de la empatía queda justificado como un mecanismo psicológico para la motivación moral. Por otro lado, el capítulo 6 es una respuesta a la explicación evolutiva de Kyle Stanford sobre nuestra experiencia de los juicios morales como si fueran externos. Defendemos que la externalización es un rasgo no solo del juicio moral, sino también del juicio de valor en general. Se sigue de ello que la evolución de la externalización no es específica del juicio moral. En segundo lugar, defendemos que los juicios de valor no se pueden separar estrictamente del nivel de las motivaciones y preferencias, que en el caso de la moralidad dependen de vínculos y demandas intersubjetivas.

Finalmente, recapitulamos las líneas maestras de tesis y las principales conclusiones alcanzadas. Destacamos también algunas de las preguntas que nos quedan por contestar y que necesitarán ser tratadas en investigación ulterior.

Resum

En el nostre dia a dia, els humans realitzem moltes conductes altruistes: ajudem els amics que ho necessiten, donem sang, fem voluntariats amb els infants del nostre barri i donem diners per a bones causes. Fem totes aquestes conductes amb la motivació d'ajudar els altres o com a resultat de la nostra consciència sobre què és el correcte. D'alguna manera, sentim que no podem sinó actuar segons els nostres judicis moral, car ens sentim especialment compromesos a actuar segons aquests. Aquest tipus de motivació que emergeix dels nostres judicis morals és el que en filosofia i psicologia moral s'ha denominat "motivació moral". Aquest és l'objecte d'estudi d'aquesta tesi.

El capítol 2 se centra en el debat sobre si la motivació moral existeix. Presentem els arguments i l'evidència empírica principals en contra de l'existència de la motivació moral; i presentem també els arguments i l'evidència empírica a favor. Per a defensar la nostra posició, ens centrem en el sentiment de culpa com a indicador del paper de la moralitat per a motivar i guiar les nostres accions.

El capítol 2 descriu la nostra psicologia moral en termes de la nostra naturalesa social, relacional. Proposem una interpretació naturalista del treball analític de Stephen Darwall al llibre *The Second Person Standpoint* (2006). Aquesta interpretació ens permet tractar alguns dels temes principals en psicologia moral, com la motivació moral, les obligacions especials, les emocions morals, els límits de la comunitat moral o l'emergència de les normes de grup.

El capítol 3 respon a la pregunta sobre com funciona la motivació moral en el context de l'amistat. En contra de Kant i Darwall, considerem que la motivació moral i les obligacions morals poden emergir d'aquest tipus de relació interpersonal. Específicament, defensem que l'amistat implica una motivació especial, a saber, motivació per la parcialitat; que l'amistat *demana* aquest tipus de motivació; i que l'amistat implica obligacions especials i implícites.

El capítol 4 se centra en la pregunta sobre evolució de la moralitat, concretada en l'origen de la motivació moral. Proposem una explicació evolutiva unitària de la moralitat que pot explicar el que considerem que són les dues principals classes de motivacions per a actuar moralment: motivació per parcialitat i motivació per imparcialitat. Reprenem la perspectiva de segona persona naturalista de la moralitat, introduïda al segon capítol, i defensem que aquesta versió naturalista pot explicar l'emergència evolutiva dels dos tipus de motivacions morals. Argumentem que la nostra proposta pot evitar les limitacions de l'explicació evolutiva de la moralitat de Michael Tomasello (2016).

Els capítols 5 i 6 exploren les implicacions de la nostra fonamentació de la motivació moral en la interacció intersubjectiva. Aquests capítols han estat publicats com a articles independents. El capítol 5 es presenta com a resposta a la posició de Jesse Prinz en contra de l'empatia. Defensem que si la moralitat i l'empatia s'entenen pròpiament, el paper de l'empatia queda justificat com a mecanisme psicològic per a la motivació moral. Per altra banda, el capítol 6 és una resposta a l'explicació evolutiva de Kyle Stanford sobre la nostra experiència dels judicis morals com si fossin externs. Defensem que l'externalització és un tret no només del judici moral, sinó del judici de valor en general. Se segueix d'això que l'evolució de l'externalització no és específica del judici moral. En segon lloc, defensem que els judicis de valor no es poden separar estrictament del nivell de les motivacions i preferències, que en el cas de la moralitat depenen de vincles i demandes intersubjectius.

Finalment, recapitulem les principals línies d'aquesta tesi i les principals conclusions assolides. Destaquem també algunes de les preguntes que ens queden per contestar i que necessitaran ser tractades en futures investigacions.

Abstract

In our daily lives, we perform many altruistic acts: we help our friends in need, we donate blood, we volunteer to prepare activities for the children in our neighborhoods, and we give money to worthy causes. We perform all those acts out of a motivation to help others and out of the clear conscience that this is the right thing to do. We somehow feel that we cannot but act according to our moral judgments, as they especially bind us. This kind of motivation which emerges from our moral judgments is what in moral philosophy and psychology has been called "moral motivation".

Chapter 1 deals with the debate about whether moral motivation really exists. We present the main arguments, and empirical evidence against the existence of moral motivation; and the arguments and empirical evidence in favor of it. To further make our point, we focus on the feeling of guilt as pointing to the actual role of morality motivating and guiding our actions.

Chapter 2 describes our moral psychology in terms of our social, relational, nature. We propose a naturalistic interpretation of Stephen Darwall's analytical project in *The Second Person Standpoint* (2006). This interpretation allows us to address some of the main issues regarding our moral psychology, such as moral motivation, special obligations, moral emotions, the limits of the moral community, or the emergence of group norms.

Chapter 3 answers the question about how our moral motivation works in the context of friendship. Contrary to Kant and Darwall, we contend that both moral motivation and moral obligation can emerge from this kind of interpersonal relationship. More specifically, we contend that friendship implies a special motivation, that is, motivation for partiality; that friendship *demands* this kind of motivation; and that friendship implies special and implicit obligations.

Chapter 4 focuses on evolution of morality, specifically on the descent of moral motivation. We propose a unitary evolutionary account of morality which can explain

what we consider it to be the two main kinds of motivations to act morally: motivation for partiality, and motivation for impartiality. We go back to the naturalistic second-personal standpoint to morality, previously introduced on the second chapter, and argue that this naturalistic version can account for the evolutionary emergence of the two kinds of moral motivations. We contend that our proposal can also avoid the shortcomings of Michael Tomasello's account of the evolution of morality (Tomasello, 2016).

Chapter 5 and 6 explore the implications that our grounding moral motivation in intersubjective interaction has in other debates. These chapters have actually been published as independent papers. Chapter 5 is presented as a reply to Jesse Prinz's position against empathy. We contend that once morality and empathy are properly understood, empathy's role in morality is vindicated as a psychological mechanism for moral motivation. Chapter 6 is a reply to Kyle Stanford's evolutionary account of our experiencing moral judgments as external. We argue that externalization is a feature not only of moral judgment, but of value judgement in general. It follows that the evolution of externalization was not specific to moral judgment. Second, we argue that value judgments cannot be strictly decoupled from the level of motivations and preferences, which in the moral case rely on intersubjective bonds and claims.

Finally, we recapitulate the main points of this dissertation. We outline some of the questions that remain to be answered and which will need to be dealt with in future research.

Contents

PROLOGUE

As I am writing this introduction (August 2019), 320 people are waiting in two different migrant rescue ships for a safe port to dock. They have been denied entry by Italy, and Malta. European politicians seem to be hardly moved by the fact that ninety percent of those on board need urgent medical assistance; neither do they seem to be moved by the fact that more than 800 people have already died this year trying to cross the Mediterranean to Europe. Sadly, this is not breaking news: avoidance of responsibility has been a frequent reaction when the countries face the challenges of immigration. We hear about it almost every day; *nihil novum sub sole*.

Intriguing and exasperating as this avoidance reaction is, I want to shift our focus onto a completely opposite reaction. Behind these images of people laying on aboard the humanitarian boats, there are people lending their blankets to those who most need them; handling the boat; and even more people anonymously funding the rescue. Furthermore, some of those who are in those migrant rescue ships are probably there because someone else helped them during their long journey through the sea; or even because their families decided to invest their money in their journeys, while they had to stay in their countries.

Far away from that boat left adrift in the middle of the Mediterranean, and way closer to my home, a group of young boys and girls will be arriving to Palma after spending two weeks of their summer looking after a bunch of children in a carefully organized summer camp. They are counselors, and they are coming back from a summer camp. They volunteer not only the two weeks of the summer camp, but also several months before, to ensure that everything is well prepared, organized and entertaining for the children. Furthermore, they spend all their Monday evenings, and some more time, to prepare activities for the children; and spend all their Saturday mornings, at least, to do those activities with them. And this is not for profit. They probably have fun doing it, but they ultimately do it for the children's sake.

Despite the terrific differences between both the migrant rescue ship, and the summer camp, there is something in common among the volunteers involved in both cases: these people act out of a motivation to help others, and the clear conscience that this is the right thing to do. They all feel that kind of urge to act that we all feel when we see that an old person is about to fall in the street; when we find a child who might have got lost; or when we hear a friend crying over the phone. In all these cases, and similar ones, we feel that we ought to act, that not acting would be faulty, that these other people can benefit from our help. Whether we finally do it or not, we still feel somehow motivated to help because helping is good, right or our duty. In moral philosophy and psychology this phenomenon is called "moral motivation". It constitutes the subject matter of this dissertation.

CHAPTER 1. INTRODUCTION: FROM GYGES TO ALADDIN

> "[On wearing the ring,] no man would keep his hands off what was not his own when he could safely take what he liked out of the market, or go into houses and lie with anyone at his pleasure, or kill or release from prison whom he would..."
>
> (Plato, *The Republic*, 2.360b)

Moral motivation has been the object of study of philosophy from ancient times. It is at the core of an old problem in moral philosophy: what motivation do we have for acting morally? Do we do it for reasons intrinsic to moral judgment, in the sense that we feel immediately motivated to do the right thing, or simply because we want to avoid the consequences that might follow from not behaving morally? In other words, do we act morally out of the strength of moral motivation or out of prudential reasons?

Plato raised this question in the second book of *Republic*. At a certain point in the dialogue, Glaucon poses this challenge through the story of the ring of Gyges (359d-360b). Glaucon tells us that one of the ancestors of Gyges used to be a shepherd in the service of the king. One day the shepherd finds a ring which has the power of making invisible those who twist it in a certain way. The shepherd uses it in his own interest: he seduces the king's wife, kills the king and takes over the kingdom. What the myth wants to show is that as soon as the shepherd knows that there will not be any negative consequences, he does not hesitate to act immorally. What Glaucon concludes is that it is not moral motivation what makes us act morally; rather the motivation to avoid the negative consequences of not doing the right thing.

Leaving the details of the myth apart, the myth presents us with a challenge: to show that moral motivation is possible and genuine, that it is possible that we act morally "for the right reason", and not out of sophisticated self-interest, out of fear of sanction, or out of a desire for retribution. Were we not to be caught for wrongdoing, would we still act morally? Is the possibility of being caught what prevents us from acting immorally, and what makes us act morally instead? Would the people onboard of the migrant rescue ships, or the counselors from the summer camp, still act morally if others could not see them? In other words, do we ever act merely out of moral motivation?

1. The immoralist challenge in moral philosophy

In our daily lives, we bump into many acts of altruism: we do favors to one another, we help our friends in need, we donate blood, we volunteer to prepare activities for the children in our neighborhood, and we give money to worthy causes. However, according to the heirs of Glaucon -- those who claim that we are psychologically egoist --, these are not really good deeds, but self-interested actions performed *as if* they were good deeds. The consequences of such a position are profound: if psychological egoism were true, what is the point of morality? Is it just a mere convention to make the most of our egoistic nature, or is it just pointless? This is what in moral philosophy has been called "the Immoralist Challenge".

Psychological egoism is the theory of human psychology which claims that everyone pursue their own self-interest. It is the position found in Hobbes (1651/2012), hence his claim that humans follow the norms of the civil society just for convenience, for their own interest. According to Hobbes, humans are not morally motivated, and even when acting according to moral norms they are motivated by self-interest. Contrary to that view, Rousseau (1762/1994) claimed that humans are indeed motivated by morality, and that they just start acting for their own interest once they live in society, and get acquainted with private property. The disagreement between the political philosophies of both Hobbes and Rousseau reflects their positions regarding psychological egoism, hence the consequences of accepting such a thesis.

As Rachels & Rachels (2015) summarize, two arguments have been typically offered in favor of psychological egoism: the argument that we always do what we most want to do; and the argument that we do what makes us feel good.

First, according to psychological egoism, we always do what we want to do; both those who act on evident self-interest and those who act altruistically are ultimately doing what they want to do. They all act according to their desires, hence on their interest. Yet this argument faces two objections. On the one hand, sometimes we act not because we *want to*, but because we *ought to*. For instance, if we had to tell a friend that their partner is cheating on them, we certainly would not *want to* do it, but we would feel that we *ought to*. In this case, our strongest desire would be to avoid the situation, yet we act against this desire. On the other hand, and assuming for the sake of the argument that we might act following our desires, whether our acting on our own desires counts as self-interest depends on the content of the desire at stake. For instance, the desire to help a friend is not self-interested, rather altruistic.

Second, according to psychological altruism, we do what makes us feel good; the fact that we feel good after performing a good action explains why we did it. Hence the real motivation of an altruistic act would be self-interest. Two objections can be raised against this argument. On the one hand, we might have several motivations simultaneously. We might help our grandparents because we want to feel good, but this does not exclude that we help them because we genuinely, and disinterestedly, care for them. On the other hand, it may also be argued that the feeling of satisfaction is a consequence of our acting morally, yet this does not mean that it works as a motive. For instance, the satisfaction I feel when I get to buy the book I desired is not the object of my desire; the book is. To the same extent, the satisfaction I feel from helping my grandparents is not the object of my desire; helping my grandparents is. Therefore, I am not acting on self-interested motives.

In conclusion, powerful as it might seem, psychological egoism has some strong objections. At bottom, the Immoralist is unable to draw the line between moral judgment and other judgments, if morality is supposed to be just a matter of social convention

(Corbí, 2003). This explains why it is not the default position in Moral Philosophy. Following this trend, this dissertation stems from the assumption that we are actually moved by moral motivation, at least some people sometimes, and that we feel somehow bound by our moral judgments. This dissertation will aim to account for how it is possible.

2. The problem of altruism in evolutionary psychology

A parallel debate about whether humans are *really* altruistic takes place in evolutionary psychology, yet it started in evolutionary biology with Darwin himself. Darwin acknowledged that his theory of evolution by means of natural selection was challenged by the altruistic behavior of both social insects (Darwin, 1859), and humans (Darwin, 1871).

The altruistic behavior of social insects remained a difficulty for Darwin. He could not account for the transmission of the altruistic behavior that only the sterile females in insect communities present. This behavior, together with other altruistic behaviors in non-human animals which were later discovered, represented a challenge to the theory of evolution by means of natural selection, and several accounts were proposed to deal with what was called "the problem of altruism". Some of the proposed accounts were group selection (Sober & Wilson, 1998), kin selection (Hamilton, 1964), or reciprocal altruism (Trivers, 1971; Wilkinson, 1990), to mention the main ones.

However, these accounts do not satisfactorily explain altruism in humans. They explain biological altruism, which consists in increasing others' reproductive fitness at a cost to one's fitness (FitzPatrick, 2016; Okasha, 2013). They focus on altruistic *behavior*, not on altruistic intentions or psychological motivation. For instance, they might explain our helping our siblings because they are part of our social group; because we share with them some genetic material; or because we assume that they will reciprocate our help later on. Yet intentions and psychological motivation in general are critical to the specific question of human altruism, as they make the difference between altruistic, and self-interested motivation. As we have seen in the discussion on psychological egoism, my helping a friend might count as self-interested if it is based on the desire to feel good afterwards; or it might count as altruistic if it is simply based on my desire to help my friend. As biological altruism does not delve into intentions, another kind of altruism needs to be introduced: psychological altruism.

Psychological altruism is defined as caring about others' welfare, or having the conscious intention to help them (FitzPatrick, 2016; Okasha, 2013). Darwin already connected this kind of altruism to human morality. According to him, humans are indeed moral creatures because they feel inclined to help others, according to their moral sense (Darwin, 1871). Darwin took the concept of moral sense from Adam Smith (1759) and David Hume (1740), thus following the sentimentalist school, according to which moral valuations show up in our emotional, or sentimental, responses. According to the followers of this school, these valuations show that humans do not just care for their self-interest, but care as well for the interests of others. They reveal benevolence, pity, and generosity.

From this position one can answer Glaucon's challenge about moral motivation: humans are indeed motivated by morality because their moral sense, being a sort of valuation sentiment, both grounds a judgment (which takes rather the form of an intuition), and drives us to action. It works this way because our evolutionary history selected it to be so. This account, that there are creatures who show a "moral sense or conscience" (Darwin, 1871) despite being in a context in which they must "struggle for existence" (Darwin, 1859), can still be challenged. It can be doubted that humans' sentimental reactions are *really* altruistic. In other words, it can be argued that the concern for the well-being of others is just a useful way to promote one's own interest, that altruism is just apparent. The debate in moral and evolutionary psychology can be simplified as a discussion between two main positions: those in favor of the existence of psychological altruism, or moral motivation; and those against it. And as the literature on this topic is extensive, I will simply state the most relevant pieces of evidence without going into any more detail than absolutely necessary.

Those against altruism typically mention some of these findings to defend that humans are not really altruistic, that is, that they are not really moved by moral motivation, or moral conscience. First, the moral hypocrisy effect: humans act immorally when they can avoid the costs of so acting by appearing morally instead (Batson, 2008). Second, the licensing effect: people act less morally after having previously appeared moral (Monin & Miller, 2001; Sachdeva, Iliev, & Medin, 2009). Third, the bystander effect: people are less likely to help when bystanders neither do it (Darley & Latane, 1968). Fourth, the effect of punishment: in economic games, people cooperate more when other participants can punish them for not doing so (Fehr & Fischbacher, 2004). Fifth, in the dictator game people do no chose "the moral option", rather "the lesser evil option" as they stop choosing the prosocial option when an intermediate alternative between the prosocial and the selfish options is offered (Levitt & List, 2007). Finally, the context-dependence effect: people are more or less likely to act morally depending on apparently irrelevant features of the context such as unconsciously seeing the image of two eyes (Bateson, Nettle, & Roberts, 2006); finding a dime in a payphone (Isen & Levin, 2017); or attending a talk about the parable of the Good Samaritan (Darley & Batson, 1973), among other examples.

In addition to reinterpreting these findings (Dill & Darwall, 2014; Sie, 2015), those in favor of the existence of psychological altruism typically mention some alternative findings to defend that humans are *really* altruistic, that is, that they are really moved by moral motivation, or moral conscience, in their concern for the wellbeing of others. First, the empathy-altruism hypothesis: people feeling empathy for a victim are more motivated to help them than those who do not (Batson, Duncan, Ackerman, Buckley, & Birch, 1981; Coke, Batson, & McDavis, 1978)[1]. Second, in the dictator game people are prosocial even though being selfish would not have any negative consequences (Forsythe, Horowitz,

[1] Dill & Darwall (2014) argue that the empathy-based altruistic motive does not count as moral motivation because it does not have the right kind of content. According to them, moral motivation must have intrinsic moral content, and hence the morally motivated person must be moved to act considering the morality of doing so. As we will argue in chapters 4 and 5, we accept empathy as part of what we call moral motivation for partiality, which is the sort of motivation the moral sense approach takes as the starting point of morality. The difficulty for the moral sense approach is with the sort of motivation for impartiality that authors of Kantian inspiration, like Darwall, take to constitute morality proper.

Savin, & Sefton, 1994; Henrich et al., 2005). Third, from a very young age, infants show prosocial tendencies and altruistic helping (Warneken & Tomasello, 2006). Finally, it is evolutionarily more plausible that humans developed a truly moral motivation than a motivation to appear moral, as the best way to appear moral is by being moral (Martínez, 2003; Rosas, 2005, 2013; Trivers, 1971). In sum, the case for psychological altruism is strong and solid.

3. The tale of Aladdin

To further defend the position of this dissertation in favor of the existence of moral motivation in humans, let me tell you another story. We have started this introduction with the myth of Gyges. Through this myth, Glaucon argues that morality is just a matter of convention, a mere instrumental good. According to him, there is no such thing as moral motivation, only sophisticated self-interest. However, it seems that moral motivation works in a different way. We do not act morally only for self-interest; and we do not avoid immoral behavior just because we do not want to be punished. Instead, we feel motivated to act morally and we experience norms as binding.

For illustration, let me tell you the story of "Aladdin and the magic lamp" from *Arabian Nights*[2]. In the Disney adaptation of the tale, Aladdin finds a magic lamp, and uses its magic power to become a prince, and hence be allowed to marry princess Jasmine. At first sight, the myth of Gyges and the tale of Aladdin seem to make the same point: if we could just act on our interest without fear of sanction, we would do it. Therefore, there is no moral motivation in humans, just self-interest, strategic thinking, and prudency.

However, as the story goes on, something relevant happens to Aladdin: once he has already won the heart of Jasmine, he does not feel satisfied. After an argument with the genie, he regrets having lied to Jasmine, making her think he was a real prince, and decides to tell her the truth. He is aware that if he tells the truth, he might lose everything he has obtained, including Jasmine's love and the sultan's confidence. Yet there is

[2] I am indebted for this example to Laura Isern Mas.

something that seems to eclipse the happiness of having his desires satisfied: the feeling of guilt.

The fact that we sometimes feel guilty shows that we are not as strategic as Glaucon pretends. Rather the opposite: even though we could act immorally and we were certain that no one would ever find out, we could never hide from ourselves. Our being aware of our wrongdoing would make us feel unease, and eventually we would probably need to repair somehow the harm done. This is the role indeed of guilt: to make an implicit valuation of wrongness, and to motivate its subject to feel responsible for their wrongdoing and to make reparations for it, such as confessing, apologizing, or undoing the consequences of the behavior (Darwall, 2006; Dill & Darwall, 2014; Strawson, 1974; Tangney, Stuewig, & Mashek, 2007; Tomasello, 2016). And the fact that guilt is a common human experience, and a central part of our psychology, demonstrates that we feel really motivated to act morally, and that we experience moral norms as binding.

One might argue that what moves Aladdin is not guilt, but just fear of the bad consequences that he might suffer if someone were to find out the truth. Yet the development of the story makes clear that what moves him is not fear of retribution: first, once lost in the mountains he regrets having been dishonest right from the start; and, second, at the end of the story he does not feel bad for not being a prince anymore, but for having acted immorally, dishonestly. More precisely, he feels bad for having acted immorally *to* Jasmine, for having lied to her. Consequently, he does not escape and tries to avoid the bad consequences of his immoral behavior; instead he finds Jasmine and apologizes to her.

This is where this dissertation departs from: we do not always act on self-interest, we can be morally motivated. We experience our moral obligations as especially binding, as powerful motives, and sometimes we act according to them truly out of moral motivation. When we fail to do so, we feel bad. When somebody else fails to do so, we blame them. Furthermore, as Aladdin shows us, both the binding force of our moral obligations, and the source of our moral motivation have a common origin: our relationships to particular others. This last point is what this dissertation will be concerned with: to articulate "the second-personal dimension of our moral psychology" and elaborate an account of its role in moral motivation.

4. The second-personal approach to morality

Our account stems from the fact that we interact emotionally and intentionally with individual others. When we interact with one another, we address each other demands, hold each other accountable for incompliance without excuse, and apologize for wrongdoing. This interaction can be explicit, through reproach, or implicit, through moral emotions, also called reactive attitudes (Strawson, 1974). If I step onto someone's foot, that person will show resentment, I will apologize, and the other will forgive me. In this second-personal interaction, we have both claimed for recognition, and we have both recognized the other's right to claim on us. It is this interaction which grounds morality, in so far as it is through these interactions that we get to feel motivated to comply with the demands that another might put on us, and get to experience them as specially binding. Hence the importance of a second-personal approach to morality.

A second-personal approach aims to shift the focus from moral judgments, and norms, to the binding force that comes from our relationships with others. Most of the accounts of our moral psychology focus on moral judgment and moral norms, and take moral emotions, and interactions as reactions to transgressions of those norms and moral judgments. Yet our moral psychology does not reduce to moral judgment; it has an essential level of prosocial motivations, and affective bonds (Cela-Conde, 1987), such as friendship or empathy. The second-personal approach to morality stems from these pre-normative forms of interaction, and considers them constitutive of our moral psychology. Consequently, it provides a continuity between human morality, and other non-human forms of prosociality (de Waal, 1996; Engelmann & Tomasello, 2017; Flack & de Waal, 2000; Jensen, 2016; Tomasello, 2016; Warneken & Tomasello, 2006).

In cognitive science and in philosophy of mind, the second-person perspective is already a prolific program of research (de Jaegher & di Paolo, 2007; Gallagher, 2005; Gomila & Pérez, 2017; Schilbach et al., 2013; Trevarthen, 1980). In ethics, its main advocate is Stephen Darwall. In *The Second-Person Standpoint* (2006), Darwall aims to ground morality in intersubjectivity. His approach is mainly analytical, and normative: he contends that the moral concepts imply second-personal interactions, but at the same time views this constitutive connection as a requirement for morality proper. He thus aims to derive the categorical imperative from such second-personal relations. His view assumes Kantian subjects, rational and free agents, as its starting point; and sees the second-personal relation as an abstract kind of relation with a justificatory role. According to Darwall, the second-personal dynamics of addressing demands and holding another accountable for incompliance are implicit in moral notions such as responsibility, obligation, and right and wrong. Hence the analytical nature of his project. Although he eventually dives into the psychology of the second-person, as he needs to specify the psychological capacities of those subjects, his project is to normatively ground morality, not to describe our moral psychology.

This dissertation is, first of all, an application of Darwall's proposal in the descriptive domain. It interprets the dynamics that Darwall describes as involved in the moral notions as real interactions that allow the emergence of our morality. Therefore, it shifts the focus from Darwall's interest in "the psychology of the second-person" to "the second -personal dimension of our moral psychology": a more descriptive, naturalistic, project which means to describe our moral psychology from a relational point of view. This naturalistic interpretation emerges as a promising way to deal with common topics in moral psychology (such as moral emotions, moral motivation or the evolution of morality), as well as common topics in moral philosophy (group norms, special duties, or the limits of the moral community).

5. Research objectives

This dissertation studies the psychology behind our moral motivation. It is structured around three questions. The first one aims at the general description of our moral psychology: can we give a unitary account of different phenomena of our moral psychology in terms of our social, relational, nature? The answer to this question works as a background from which we can deal specifically with moral motivation. The second question focuses on the applied approach to moral psychology: how does our moral psychology work in moral actions involving our friends? From the answer to this question, an original account of moral motivation, and our experience of moral obligations as binding follows. Finally, the third question dives into the evolutionary plausibility of this account of moral motivation: how did our moral motivation evolve?

This dissertation is intended, then, as an addition to debates in different areas. I delve mainly into moral psychology and descriptive evolutionary ethics; and marginally into meta-ethics. I do not deal with normative ethics because my project is not to justify or undermine any normative ethical claims or theories. I stick to the descriptive approach: I aim to describe what human moral psychology is and why it came to be that way. We are not in the business of prescribing how it should be.

Apart from its theoretical and philosophical aims, this dissertation has implications in both experimental psychology, and ethics. It has implications in experimental psychology because it puts forth some predictions which can be tested empirically. Hence, it opens the door to the design of future studies. On the other hand, it has implications in ethics. Determining how our moral psychology works, and how it has been shaped by evolution is not only important from a purely descriptive perspective, but also from the point of view of ethics. If we find out that our moral psychology hardly fits with what normative ethics requires from us, some "eternal" moral obligations should be reconsidered. To require something from someone we need to know what that agent can actually do: nothing newer than the old "ought implies can" principle. It seems that moral psychologists can really help moral philosophers and ethicists in this sense. This dissertation can be seen as an attempt to build a bridge between both philosophers and psychologists; and between prescriptive and descriptive ethics.

6. Structure of the dissertation

As I said, this dissertation is structured around three questions. Since every chapter is meant to be self-sufficient, each of them is devoted to a question. However, they are all part of the larger project of emphasizing the second-personal dimension of our moral psychology, and hence they all share the same core ideas. To preserve this intended self-sufficiency, these core ideas are repeated in each chapter, when needed, and exposed in line with the argumentative proposal at stake.

Chapter 2 addresses the question of the description of our moral psychology in terms of our social, relational, nature. We propose a naturalistic interpretation of Stephen Darwall's analytical project in *The Second Person Standpoint* (2006). This interpretation allows us to address some of the main issues about our moral psychology. First, we explain why moral norms motivate us; namely, because of the second-personal relations that we establish with others. Second, we articulate how intersubjective interactions take place effectively; grounding duties to particular other subjects, and being related to distinctive moral emotions. Third, we address the question of the limits of the moral community, proposing that it comprises all agents capable of second-personal interactions. Finally, we explain the emergence of community norms through intersubjective interaction.

Chapter 3 answers the question about how our moral motivation works in a specific context: moral obligations involving our friends. We ask ourselves whether the love that we feel for our friends plays any role in either our moral motivation to act towards them; or in our moral obligations towards them, that is, in our special duties. We chose this topic because it is a kind of touchstone for the Kantian approach for its emphasis on impartiality –something incompatible with acting out of love. Contrary to the Kantian approach, we contend that love plays an essential role in both our moral motivations and obligations. To articulate our proposal, we first spell out the Kantian position, and next present Darwall's second-personal version of it. According to Darwall, love is not necessary or sufficient in moral motivation, neither in moral obligation, as they are both grounded in the second-personal dynamics of accountability, which is inspired by contractualism. We raise three difficulties for Darwall's proposal: it is psychologically inaccurate, undesirable in practice, and hardly conceivable in theory. To put it another way, we argue that the Kantian cold morality Darwall prescribes sets, not just a very stringent standard of morality, but rather an undesirable one. To avoid these problems we propose to view the kind of interaction that Darwall emphasizes, not as one between partners to a contract, but as a source of mutual commitments through affiliative relationships. This drives us to articulate the sort of psychological capabilities for interpersonal interaction that such interpersonal relationship requires –what we call a second-personal approach to our moral psychology. We contend that both moral motivation and moral obligation come from our interpersonal relations with particular others, that is, from our second-personal relations with others. From this psychological version of the second-person standpoint, the three problems raised to Darwall's position are clarified, and the role of love is emphasized in both our moral motivation, and our moral obligations towards friends.

Chapter 4 focuses on the question about the evolution of morality. We propose a unitary evolutionary account of morality which can explain what we consider it to be the two main motivations to act morally: a motivation for partiality, such as the one love grounds, and a motivation for impartiality, which full-blown morality requires. We describe these two kinds of motivation and argue for the role of both of them in the psychology of moral motivation. We depart from a naturalistic second-personal standpoint to morality, which we view as presupposed in Stephen Darwall's theory, and argue that this naturalistic version has the potential to account for both motivations in the evolution of morality. Furthermore, it can also avoid the shortcomings of Michael Tomasello's account of the evolution of morality (2016). We discuss some evolutionary corollaries which follow from our proposal.

Finally, Chapter 5 and Chapter 6 explore the implications of this dissertation in other debates. These chapters have actually been published as independent papers. Chapter 5 is presented as a reply to Jesse Prinz's position against empathy's role in morality (Prinz, 2011). First, we show that even conceding Prinz his notions of empathy and moral competence, empathy still plays a role in moral competence. Secondly, we argue that moral competence does not reduce to moral judgment, as Prinz assumes, but that motivation is also relevant. Third, we reject Prinz's notion of empathy because it is too restrictive, in requiring emotional matching. We conclude that once morality and empathy are properly understood, empathy's role in morality is vindicated. Morality does not reduce to a form of rational judgment, but it necessarily presupposes prosocial preferences and motivation and sensitivity to intersubjective demands.

Chapter 6 is a reply to Kyle Stanford's evolutionary account of our experiencing moral judgments as external (Stanford, 2018). We argue that externalization is a feature not only of moral judgment, but of value judgement in general. It follows that the evolution of externalization was not specific to moral judgment. Second, we argue that value judgments cannot be decoupled from the level of motivations and preferences, which in the moral case rely on intersubjective bonds and claims.

In the final Conclusions, we recapitulate what we have achieved through this dissertation and outline some of the questions that remain to be properly answered and which need to be dealt with in future research.

CHAPTER 2. NATURALIZING DARWALL'S SECOND-PERSON STANDPOINT

"For the *I* of the primary word *I-Thou* is a different *I* from
that of the primary word *I-It*"

(Martin Buber, *I And Thou*)

1. Intersubjectivity and the grounds of morality

There have been several attempts to ground morality in the way we relate to others. First attempts in this line can be found in the Idealist school. In reaction to Kant's abstract deontology, authors such as Fichte (1797/2000), or Hegel (1807/1979), formulated the idea that mutual recognition sets up the subject as a subject first, and as a subject of rights, next. To become a self, one needs to confront other selves. Thus, their focus was on the process of constitution of a subject, trying to spot the relevant interactions to this extent. In the 20th century, in contrast, the most salient attempts at developing this intersubjective approach, such as Ricoeur's (1954) or Lévinas' (1969), were of phenomenological character. They contended that some basic human experiences, such as sympathy or compassion, or even eye contact, already involve a normative dimension. However, in this phenomenological tradition the 'other' is not conceptualized as another subject, but as part of what appears in my conscious experience. Hence, reciprocity and interaction are not properly relational, but just considered as they are experienced (Gomila, 2001a). Besides, and more important, this tradition contends that these experiences ensure somehow the universal recognition of humankind, maybe as a compensatory reaction towards the hard moral experience of the 20th century. However, that very same moral experience shows us that some people do not feel sympathy or pity towards others and their harm, and some can even cause that harm (Gomila, 2008). In other words, intersubjectivity cannot ensure subjects who are morally good, and sensitive to others' demands and needs; there is no ultimate and transcendental agency which ensures this hope.

Stephen Darwall's project of a 'second-personal morality' (Darwall, 2006, 2013a, 2013b) is something different, even if it also takes intersubjective interaction as a touchstone. To begin with, it is analytical and normative, rather than descriptive or explanatory. It aims to show that moral notions, such as duty and obligation, constitutively imply our second-personal interaction with others. In a second-personal interaction a subject addresses a claim or demand to another, who can recognize the claim or demand as valid or not. And through this interactive dynamics of claims, recognitions, mutual demands and reasons, both subjects hold each other accountable. According to Darwall, it is this holding each other accountable that is implicit in the moral notions, such as respect or dignity. Similarly, this second-personal network of accountability and recognition justifies the Kantian formal principle of normative universalization.

Darwall's proposal does not delve into the nature of our actual interpersonal relationships. As we have said, his proposal is normative at heart. He draws from rational and free agents in a kind of interaction, second-personal interaction, which does not actually need to take place. However, if we draw from flesh and blood subjects in their particular interpersonal relationships, that is, subjects who bond with each other, we can give a more accurate account of our moral psychology. In this paper, we make explicit the naturalistic dimension presupposed by Darwall's theory, and use it to provide an account of our moral psychology.

In section 2 we describe Darwall's analytical project of connecting intersubjectivity and morality. In section 3, we contend that this connection presupposes a naturalistic articulation. This naturalistic approach allows us to address some of the main questions in moral psychology, and meta-ethics. We provide a naturalistic account of the motivational power of moral judgments, in section 4, and of our special obligations to particular subjects, in section 5; both based on intersubjective interactions. In section 6, we show how intersubjective interactions are already to be found in moral emotions. In section 7, we tackle the question of the limits of the moral community, proposing that it comprises all agents capable of second-personal interactions. And, finally, in section 8, we explain the emergence of community norms through intersubjective interaction. Needless to say, our project is not to propose a re-interpretation of Darwall's work, which is neither descriptive nor naturalistic, but rather an original account of our moral psychology inspired in his work.

2. The second-person standpoint as a conceptual analysis

Darwall defines the second-person standpoint as "the perspective you and I take up when we make and acknowledge claims on one another's conduct and will" (Darwall, 2006, p.3); the perspective we take in the practices of holding each other accountable and responding to those claims. According to Darwall, these second-personal practices are relevant for morality because moral notions involve second-personal notions, and because the grounds of moral motivation lie in the second-personal relationship. Moral notions do not stand in a rational heaven, but presuppose those second-personal practices among moral subjects.

To account for the second-person standpoint, Darwall proposes the following situation as the paradigmatic instance of a second-personal interaction. The interaction starts when one person steps on another's foot. Hereafter, the person who steps on the other's foot will be the "transgressor", and the person whose foot is stepped onto will be the "victim". Being persons, and hence having equal dignity, they both have the authority to demand a certain treatment of each other. Furthermore, as their relation is governed by what Darwall calls "reciprocal recognition" (Darwall, 2006, p.48), they both recognize each other and can address demands to each other. Accordingly, the victim demands the transgressor to move his foot; and the transgressor knows and feels that they ought to accept and respect the victim's claim. This feeling comes from the second-personal nature of the relationship at issue: it is a relationship of reciprocal accountability through which they address demands to each other, and hold one another responsible for compliance. At the same time, the victim reacts to the transgressor's reaction, accepts their stepping behind, and the relationship is reestablished. In this way, the problem of moral motivation finds another, more promising answer: the purported authority of morality, i.e. its motivational power, derives from this recognition of others as sources of obligation.

Notice that Darwall's account is analytical in the first place. It does not describe what is going on in cases of harming one another, but claims to unpack the content of the notions that characterize morality. In other words, morality constitutively requires that these patterns of mutual recognition, of addressing claims and honoring them, take place. In this way, the account also becomes normative in that at the same time it sets the standards for real world human interactions to count as properly moral. Mutual respect becomes mandatory for moral agents as long as it is implicit in the very notion of morality.

This analytical and normative stance entails that Darwall does not want to accept the consequence that moral obligation derives from mutual demands. According to him, stepping on the victim's foot is wrong even if the victim does not protest. In fact, the transgressor's feeling that they ought to respect their victim's claim and move their foot comes from the transgressor's knowing that they could justifiably be held accountable for incompliance, even by themselves in their own conscience, as if they adopted a second-person standpoint towards themselves. If the transgressor did not move their foot, they would be accountable to their victim's claim for respect of their dignity as a person, even if the victim did not make it. Consequently, the transgressor accepts the victim's right to claim, and reacts to it by moving their foot.

In this pattern of interaction, several assumptions and concepts are in play, according to Darwall. To make them clear, we will start with the assumptions; and move next to the concepts, which constitute an "interdefinable circle" (Darwall, 2006, p.12) where each one implies all the rest. Hence the analytical nature of the approach. Let us disentangle each assumption and concept one after the other, and see how they relate.

In the practice of holding accountable, i.e. of giving and asking for reasons, agents involved in an interaction assume that they both have: (1) a right to make claims, to demand respect for those claims, and to resist the demands of the other; (2) a second-personal authority to make demands or claims, and to hold the other accountable for non-compliance without excuse; and, (3) a dignity which must be respected and which cannot be violated by any claim. As this characterization makes clear, Darwall's second-personal standpoint is normative, rather than descriptive. It specifies a desideratum, rather than describes how things always happen. The circle of concepts specifies the way in which particular interactions should take place to qualify as properly moral.

One of those concepts is second-personal authority. Second-personal authority is the authority that a moral subject has to address claims, and demands to other subjects. For the addressee, the claim creates a distinctive reason for compliance. In Darwall's paradigmatic example, both the transgressor and the victim have authority to address demands to each other. Furthermore, the victim's demand that the transgressor move their foot makes the transgressor responsible for complying, since the transgressor recognizes the practical authority of their victim. This practical authority is presupposed when an addresser claims or demands something of an addressee, and this addressee recognizes the addresser's right to so claim. Therefore, the concept of authority is necessarily tied to other second-personal concepts. First, second-personal authority is second-personal because it assumes that it is addressed to particular subjects in interaction. Second, it entails second-personal competence, which means that "whenever second-personal address asserts or presupposes differential authority, it must assume also that this authority is acceptable to its addressee simply as a free and rational agent" (Darwall, 2006, p.22). Besides, the addresser must also assume the addressee's capacity of free self-determination to accept internally the authoritative demand, and decide whether or not to respond to it. Finally, the notion of second-personal authority involves necessarily the notion of responsibility or accountability. The authority to demand implies not just a reason for the addressee to comply, but also their being responsible for doing so and their accepting the possibility of being held accountable for non-compliance without excuse.

As for second-personal responsibility, it "concerns how, in light of what someone has done, she is to be related to, that is, regarded and addressed (including herself) within the second-personal relationship we stand in as members of the moral community" (Darwall, 2006, p.69). We see this notion graphically illustrated in daily situations such as those in which a caretaker scolds a child for something they has just done. In most of these cases, the caretaker points at a drawing in the wall, or a messy table, and yells at the child "look what you have done". Somehow, this caretaker is trying to make explicit the relation of the child to what they has done, and to hold them accountable for it. Indeed, Darwall understands responsibility as accountability; in other words, we are responsible for what a member of the moral community can hold us accountable for doing (again, even if nobody never does).

What invests agents with authority is the fact that they can address claims to each other. Such claims provide reasons both explicitly through speech acts, such as reproaches, excuses, or requests; or implicitly through reactive attitudes such as resentment, indignation or anger. For instance, the person whose foot is stepped onto can either verbally ask the transgressor to move their foot, and hence address explicitly a reason through a speech act; or make a gesture of protest showing their disapproval, expressing resentment at the transgressor's action, and hence addressing the reason implicitly through a reactive attitude. These reactive attitudes are forms of address that directly appeal to the addressee's goodwill and that hold him accountable for compliance.

In these dynamics of second-personal interaction, as spelled out by Darwall, what is given through a demand is a second-personal reason. The victim's resentment at the transgressor's stepping onto their foot counts as a reason for the transgressor to move their foot. This reason is second-personal because it has the following features. First, it is an agent-relative practical reason; that is, it is a reason for acting "whose validity depends on presupposed authority and accountability relations between persons and, therefore, on the possibility of the reason's being addressed person-to-person" (Darwall, 2006, p.8). Second, it aims at motivating the other's will through the agent's own self-determining choice. Accordingly, it is not a kind of coercion, but an internal acceptance of an authoritative demand. In Darwall's example, the victim seeks compliance after recognition, instead of mere obedience from the transgressor. Third, and as a result of being part of the dynamics of the second-person standpoint, second-personal reasons presuppose that both agents have equally second-personal authority, competence, and responsibility as free and rational agents, and that they can exchange their positions as addresser and addressee.

In summary, according to Darwall the validity of second-personal reasons depends on the authority and accountability relations between addressees and addressers, who can exchange their roles in their interaction; and on the ability of the participants to self-determine themselves freely by acknowledging the authority of the other agents they interact with. However, Darwall avoids the conclusion that claims are justified if they are addressed, or accepted. According to Darwall, the justification of those claims is established independently, in so far as they are universalizable, à la Kant. That is why he insists that the demands can be internally recognized, as if the second-person standpoint was internalized. On the contrary, a naturalistic approach assumes that the sort of impersonal point of view required to establish the universalizability of claims the analytical approach requires does not exist. As a consequence, the legitimacy of moral norms and practices of mutual recognition has to be viewed as grounded in the dynamics of intersubjective demands.

3. A naturalistic approach to the second-person

So far we have introduced the basics of Darwall's analytical project. In what follows, we try to develop it into a naturalistic framework, as we think it provides useful elements for an account of our moral psychology. As we have already remarked, Darwall's approach is presented as analytical. Yet, as a matter of fact, it turns out to be normative, as it assumes rational and free agents, and takes the normative dimension as universal and independent of the particular claims agents address to each other. Darwall's model is inspired by contractualism at heart: rational subjects recognize each other as such, giving rise to mutual respect, reciprocal claiming, and deals. All rational subjects are interchangeable, and they have internalized this second-personal standpoint: their moral conscience results from this self-assessment according to already legitimate demands. However, the sort of second-personal interaction that Darwall describes also invites another way to develop it: as a naturalistic account of how such dynamics of claims and respect for them takes place, and how the remarkable kind of moral agents can emerge in the first place; how such subjects are constituted by the sort of intersubjective interactions described; and to what extent we humans can approach such normative ideal of rationality.

For instance, Darwall's free and rational agents need to be endowed with a full set of psychological capacities. They have to be linguistic beings to verbally address and receive claims. They also have to be emotional beings, to address claims implicitly through their reactive attitudes. They also require some degree of self-control and self-regulation, to behave in a self-initiated way. And they need to understand others, through some form of psychological attribution, to recognize others' intentions, plans, and emotions, and respond to them. Maybe also some kind of empathy, compassion, or sympathy needs to be presupposed.

Thus, although Darwall's project is formulated as analytical, it also calls for a naturalistic project, which accounts for flesh and blood subjects who are somehow sensitive to others' demands, and who need affectively bonding with others. Hence, in our view Darwall's proposal requires a naturalistic counterpart, and offers the seeds for it, even if Darwall is not interested in such a project. Similarly, an appropriate naturalistic project can benefit from his characterization of morality as intrinsically second-personal, to articulate the way in which our moral psychology is shaped by second-person interaction.

In this way, Darwall's second person standpoint presupposes a more basic notion of second personal interaction, as the way we come to interact with particular others, before and beyond accountability –an approach we have tried to independently develop (Gomila, 2002, 2008, 2015). Instead of characterizing the second-person perspective as intrinsically moral, the naturalistic approach focuses on how the intersubjective structure of recognition emerges in interaction. In this view, the second-person perspective is the way in which we attribute mutually, and implicitly, expressive mental states, such as intentions and emotions, to those with whom we interact face to face (Gomila, 2001a, 2002, 2015). It is our spontaneous way to make sense and adjust to others' behavior in face to face interactions. From this point of view, the second-person perspective characterizes the psychological competence which ensures mutual understanding in intersubjective interaction (Gomila, 2008).

In this naturalistic project, morality is still grounded in the second-person perspective because morality requires our ability to interact with others in an intersubjective way (Gomila, 2008). We need to attribute intentions, keep track of epistemic states and recognize emotional expressions, if we are to make sense of the claims and demands that others may address to us, and respond properly. Think, for instance, of Strawson's reactive attitudes, which Darwall takes as a case of implicit second-personal claim addressed to another. To react with resentment to another's deeds, we need to see their behavior as intentional in the first place, and with a particular intention, given the context. Thus, Darwall's normative project presupposes our naturalistic project of the second-person as the perspective of intentional interaction in real time. Furthermore, in Darwall's project agents need not simply to attribute or recognize an emotion in their interactive party in a distanced, off-line, third personal way; but to react emotionally and intentionally to them in an engaged, online, reciprocally contingent, second-personal way (Gomila, 2002). Darwall does not consider in detail the complexities of psychological attributions, and just assumes the view of simulation theory (Goldman, 1992b; Gordon, 1992).

This naturalistic understanding of the second-person standpoint can account for our moral psychology. In what follows, we use our naturalization of Darwall's project to put forward five essential issues in moral psychology: the motivational power of moral judgments; special obligations we recognize towards particular subjects; the intersubjectivity involved in moral emotions; the limits of the moral community; and the emergence of moral group norms.

4. The motivational power of moral judgments

One of the topics of moral psychology which can be addressed from this naturalized view of the second-person standpoint of morality is the phenomenon of moral motivation.

Moral motivation is the phenomenon by which we feel somehow motivated to act in accordance with our moral judgments (Rosati, 2016). For instance, it is what makes us feel obliged to help friends in need. It is not just that we judge that it is morally correct to help them by applying some general norm to the situation in question. For if moral reasoning were like that, our moral judgment would be similar to our judgment to drive on the left side when we are in Great Britain: this is the correct thing to do in the situation, but relative to a context. It would be a sort of prudential judgment, based on the willingness to comply with the established norms and on the fear of the negative consequences of not doing so. But moral judgments are different. Moral judgments motivate in a distinctive way: we feel obliged to comply with them. This sense of obligation can involve several aspects: a bodily sense of urgency, an anticipation of shame at the thought of failing to comply, remorse and lower self-esteem if failure did happen. Morality contributes to our identities (Riis, Simmons, & Goodwin, 2008; Strohminger & Nichols, 2014, 2015; Tobia, 2015).

For instance, in Victor Hugo's *Les Misérables*, when Marius recognizes that he ought to stay at the barricade and fight with his colleagues instead of running after his beloved Cosette, he does not just recognize his duty, he also feels that he cannot but comply with it. He feels motivated to do what he thinks his duty is. Not only does he judge that fighting is his duty, but he also feels unease at the idea of running after Cosette. He anticipates a sense of cowardice and treachery if he were to escape from the fight to go after his beloved. His decision to stay at the barricade is not just a deliberate decision out of a prudential calculation of pros and cons, of gains and losses. It is not the choice of the most optimal benefits. It is neither the result of the application of a general principle to a particular case. It is the consequence of his feeling the urge to stay, to comply with his duty because failing to do so would amount to reveal an evil moral identity.

Moral judgments, then, have this dual character. On the one hand, they are not just a matter of taste, or of personal inclination; they are truth-apt, as moral cognitivists would emphasize. For instance, fighting at the barricades is felt by Marius as the right action for anyone in a similar position. It seems to be something externally imposed, a truth to be recognized, maybe learned. But at the same time, moral judgments are experienced psychologically in a way that descriptive statements do not: they involve emotions about oneself and one's sense of self. They are strongly related to motivation, as moral non-cognitivists notice. They are experienced as a subjective commitment. Thus, before the meta-ethical debate about their ontology, what we can claim from our naturalistic view is that moral judgments are psychologically experienced both as objective and subjective (Isern-Mas & Gomila, 2018).

From the second-person standpoint, the motivational power of moral judgments can be explained by their second-personal character. As we have mentioned, according to Darwall, we feel obliged to follow our moral judgments because they imply the recognition of the legitimacy of the claim of another agent. Remember the case of the person whose feet the transgressor stepped onto. The victim has a second-personal authority to hold the transgressor accountable if they rejected to move their foot. By respecting the addresser's claim right and second-personal authority, the addressee, i.e. the transgressor, is responsible for compliance and must be prepared to be held accountable if they does not comply. And it is this knowing that they could justifiably be held accountable which motivates the transgressor to comply with the demands that other members can legitimately address to them, i.e., their moral obligations, according to Darwall.

Therefore, moral judgment has motivational power on us because it is essentially interpersonal, intersubjective or, in Darwall's terms, second-personal. For instance, I feel, and know, that I ought to help my friends in need, because this is what them, or any other members of the moral community including myself, could demand me to do; or could hold me accountable for not doing. Likewise, I ought not to mistreat my partner, because if I did it, he would have the right to hold me accountable for doing it, and I should blame myself too. Accordingly, the motivational power of moral judgment does not come from our being aware of the moral law through moral deliberation, *à la* Kant, or from a calculation of consequences, as utilitarianism prescribes. It comes from the motivational nature of second-personal claims and reasons; which motivate us because they come from a recognized authority, so that they become internalized. In fact, moral obligations are defined as "what those to whom we are morally responsible have the authority to demand that we do" (Darwall, 2006, p.14), or "what the moral community can demand (and what no one has the right not to do)" (Darwall, 2006, p.20). Therefore, it is because we assume as our own (at least some of) others' moral demands: those that we honor and find justified. Therefore, the motivational power of morality derives from our receptiveness to the demands of those with whom we interact (Isern-Mas & Gomila, 2018).

All this structure need not to be explicit, as Darwall himself recognizes. The Strawsonian reactive attitudes (Strawson, 1974) implicitly involve this web of reciprocal expectations. For instance, my knowing that my friends could hold me accountable for not helping them is manifested in my feelings of guilt, which is just "to feel as it one has the requisite capacity and standing to be addressed as responsible" (Darwall, 2006, p.71). This emotion-laden interaction is what makes us feel bound by others' demands on us, and eventually by the moral norms which will emerge from those interactions. Indeed, according to Carla Bagnoli, the "apparent inescapability of moral norms and the specific kind of authority that they have in our minds" is due to moral emotions (2006, p.8). Therefore, it is through emotions that we feel motivated to act according to our moral judgments.

> If we had no moral sensibility we would have only extrinsic motives to enforce moral norms, such as sanctions and incentives, fear of punishment and expectation of reward. We are able to undertake morality as a subjective motive because we are capable of moral sensibility. (Bagnoli, 2006, p.13)

The naturalistic twist can offer an account of why we are so susceptible to each other's demands. The answer has to do with the fact that we are a social or, more precisely, an "ultra-social" species (Tomasello, 1999, p.59). Our evolutionary origins make us feel motivated to bond with others and to take their interests and needs into account (Cheney & Seyfarth, 2008; Seyfarth & Cheney, 2012; Tomasello, 2016). Within this evolutionary path of interpersonal dependencies, morality seems to have emerged as a game of reciprocity requests that are recognized and self-imposed, instead of imposed through coercion, or fear of punishment. Evolution made us pro-social in the first place; and morally motivated afterwards.

5. Special obligations towards particular subjects

The naturalistic approach to the second-person standpoint allows us to capture the complexity and particularity of our real interactive, and intersubjective, relationships. As already mentioned, Darwall's analytical approach avoids this line of reasoning. But a naturalistic approach of the second-person standpoint makes clear that the kind of relationship that binds subjects is not an abstract one that holds equally with any member of the moral community. Rather, it is a particular, emotionally-loaded, relationship which is established with specific persons during our lifetime; and which invests each other with variable degrees of authority to yield particular demands, depending on the situation and the agents involved. For instance, our friends' demands bind us in a different way than our neighbors' or any strangers' ones do.

As Wallace notices, "those who are implicated in a nexus of relational normativity possess a kind of practical authority over the relevant normative relation that uninvolved third parties lack" (Wallace, 2007, p.29). For instance, the victim of a moral transgression has both a "privileged authority to complain" (Wallace, 2007, p.29), and the possibility to consent a behavior against her dignity that would be otherwise considered a transgression. Consequently, "the person who is wronged by you has a privileged basis for complaint against you, an objection to your conduct that is not shared by mere observers to what was done" (Wallace, 2007, p.29). Aiming at our moral psychology, this view seems correct.

Darwall makes sense of this experience through the notions of "obligations of loving relationship" (2016, p.172), or "duties of relationship" (2016, p.177). These are moral duties that are shaped by the specific circumstances where they take place, and which are addressed to "a-person-who-happens-to-stand-in-that-specific-putatively-normative-relation" (2006, p.270). He seems to think, for example, of intergenerational duties: all sons and daughters have certain duties, in virtue of being sons and daugthers, towards their parents. These duties are still moral, because they pass Kant's test of universalizability and impartiality. We could not conceive a world where victims did not have a special authority towards transgressors; where friends did not have a special authority towards other friends; and where social causes were not given more weight than personally romantic desires. Therefore, Darwall acknowledges the special authority of subjects standing in particular relationships.

Nevertheless, Darwall's point is rather that blame is assessed not from the victim's point of view, but from anyone's, from an impartial standpoint. In his own words, "although resentment is an attitude that can intelligibly be felt only from a victim's individual standpoint, or from one that identifies with it, blame can be felt from anyone's standpoint; it entails the representative authority of the moral community" (2018, p.814). According to Darwall, it is critical that the claims and demands that are addressed in a particular circumstance are justified, and sanctioned from anyone's point of view, and this entails that its legitimacy is independent of anyone's particular connection to the situation. Darwall finds inspiration in Adam Smith's (1759) notion of an impartial spectator. Through this notion he connects the dynamics of particular claims and the Kantian procedural requirement of impartiality as the validity criterion for moral claims. Therefore, even our special obligations towards our friends should be considered as such by anyone who was in our position in relation to them. Hence, as explained in section 1, Darwall's project is Kantian at bottom. In Darwall's morality, the demands that we address to each other must ultimately be ones that could be endorsed from a "perspective that we can all share as free (second-personally competent) and rational" (Darwall, 2006, p.276); they are grounded in our "common authority to make claims on each other" (Darwall, 2006, p.274). In Wallace's terms, Darwall's picture of morality is "one on which normative principles get traced in the end to a kind of (hypothetical) collective selflegislation, whereby we make principles normative for ourselves by imposing them on ourselves from a common point of view" (Wallace, 2007, p.32).

However, this common point of view is difficult to sustain within a naturalistic framework. On the one hand, there may not be a unique way to generalize other's perspectives. There are many situations in life where we relate to others in ways that are not institutionalized, which may be diverse and new, and which may give rise to blameworthy actions without excuse. In these situations we relate to those who harmed us from the point of view of someone harmed by them, because of the particular relationship held with them. These actions matter first of all to us, as people involved in that particular circumstance, and provide us with a distinctive authority over those who harmed us (Corbí, 2005). Saying that we could find a description of the situation that would allow us to generalize the demand to any agent in the same circumstances misses a psychological point: it is those involved in the particular relationship that address claims and recognize duties to each other; and it is also them who have a "unique position to alter the normative relations at issue" (Wallace, 2007, p.29) by consenting the kind of behavior that would be otherwise prohibited for the sake of his position as bearer of the violated right.

From this point of view, the practice of blaming takes place first of all within a particular relationship, where participants hold each other accountable depending on their specific stories of relation and affection. In this view, subjects feel bound not by an abstract relationship with any member of the moral community, but by a particular and emotional relationship which is established between specific persons. It is this particular relationship that determines the scope of accountability, and the extension and content of blame. Furthermore, within particular relationships it may also happen that a claim is not recognized as valid, or that an excuse is generated to prevent such a request.

This is nicely expressed by Antoine de Saint-Exupéry in his book *The Little Prince.* Talking about his rose to other roses, he acknowledges that *his* rose has special claims over him, which other roses do not have. Specifically, in chapter 21 he says:

> But in herself alone she [*the rose*] is more important than all the hundreds of you other roses: because it is she that I have watered; because it is she that I have put under the glass globe; because it is she that I have sheltered behind the screen; because it is for her that I have killed the caterpillars (expect the two or three that we saved to become butterflies); because it is she that I have listened to, when she grumbled, or boasted, or ever sometime she said nothing. Because she is my rose.

The reason for the distinctive authority that the rose has over the Little Prince is, as he notices, that they have a special relationship; they are friends. Consequently, he has some special duties towards his rose, by virtue of being *that* rose.

The feeling of being bound by the moral norm still comes from the fact that we can be held accountable for non-compliance without excuse. But now there is a difference between the persons that set demands on us as members of the moral community, and those who do it as particular persons that stand in particular relationship with us. Our feeling of being bound by the moral law is enhanced because of the special authority that different persons have on us, due to our relationships with them. For instance, the claim that comes from the person whose foot I stepped into binds me especially because I have a particular relationship with that person as the victim of my transgression. Remarkably, each agent may be part of multiple such relationships within a community, constituting a network of interpersonal links. As we will see in section 8, it is this web that helps to explain how a community's normative common code can emerge and be shared.

6. Moral emotions

Conceived in the naturalized manner we propose, the second-person standpoint is specially linked to moral emotions. First, because moral emotions have a role in our communication of demands, as we explained in section 2. We can react to what another did to us by expressing emotions whose content implicitly involves an appraisal of the particular episode of relation, and in this way, we demand recognition from her. The achievement, or failure, of recognition triggers also specific emotions. Second, moral emotions also have a role in the way we establish and sustain relations to others. Moral subjects (or persons) experience the need to bond affectively and emotionally with particular others, for instance through relations of friendship, trust or love; hence, we need to interact second-personally with others.

Darwall focuses especially on the first point: moral emotions as a way to implicitly address demands or react to other's demands. He relies on the way Strawson characterized them as "reactive attitudes" (Strawson, 1974), to point out that they amount to implicit forms of blaming that assume that agents are accountable for their actions. Reactive attitudes address implicit demands, whereas explicit demands require verbal expression. In Darwall's account, a reactive attitude is a form of communication which takes place in interpersonal interaction, which is elicited as a response to a person's behavior, and which seeks to reestablish the recognition and reciprocal respect that participants owe to each other as members of a community of mutually responsible agents. Thus, they entail that both participants must recognize themselves and each other as fully morally responsible agents who are able to participate in adult relationships.

For instance, if I am roaming the streets and suddenly someone pushes me away and does not apologize, I will feel unrecognized as an agent who deserves apologies, and I will probably react to this with indignation. Upon noticing my implicit claim and recognizing it, the agent is expected to repair the situation by expressing regret, or even helping me. This pattern of psychological interaction implicitly involves a basic level of normativity, that is, it demands a right attitude from the other towards me, and from me towards the other.

As explained in section 3, our naturalistic approach to the second-person standpoint characterizes the psychological competence which ensures understanding, and hence mediates second personal interactions. We interact with particular others by implicitly and mutually attributing mental states in an online, emotionally and intentionally engaged way (Gomila, 2001a, 2002, 2015). Moral emotions are an example of this kind of spontaneous, implicit and online interaction. From this point of view, moral emotions constitute an intersubjective means of implicitly addressing claims to others, as Darwall and Strawson notice. But as moral emotions develop in time, they also constitute a story of the relationship, which explains the web of affiliations and preferences from which particular duties derive.

Secondly, moral emotions build bonding relationships with others. Not only are moral emotions a way to communicate a demand, they also give rise to affiliative attachments. I may feel guilty for not having paid enough attention to someone who made me a favor before; I may feel humiliated by somebody that time after time ignores my opinions; or I may feel resentful at somebody's reluctant way to excuse their transgression, because when the roles were reversed, I felt fully accountable for what I did. Through this kind of sequences of interactions we generate personal preferences, which may become moral norms. Again, a naturalistic approach goes beyond Darwall's.

As an illustration of this twofold role of moral emotions, consider the following example[3]. Imagine I promised my friend Patrick that I would have dinner with him tonight. On my way to Patrick's, I see myself involved in a car accident where I am the only one who can help the victim. Although I am still morally obliged to keep my promise, I am also morally obliged to help another in need, especially when nobody else can. So I help that person while assuming that Patrick will feel resented; and I will probably feel guilty, for not keeping my promise. Remarkably, both reactive attitudes are not justified, according to Darwall. As what I did was morally correct, given the circumstances, I am not blameworthy, and hence Patrick's remorse and my guilt turn out to be misplaced. However, when we take into account the second role of moral emotions, we realize that what I did damaged to some extent my friendship with Patrick. In fact, it would be surprising if I did not feel guilty for failing to keep my promise to him. A similar and historical example of these fitting although unjustified emotional responses is to be found in Primo Levi's and other survivors' experiences of shame, and guilt after their liberation from Auschwitz, as if their survival was a treason to their dead fellows (Levi, 1986/2017). Although in this case it is an open question whether these moral emotions are actually

[3] We are indebted for a similar example to Stephen Darwall.

justified (Corbí, 2005), what these examples show is rather that the way we are connected to other people influences which actions and omissions are viewed as cause for a claim of proper respect and recognition; and which emotional reactions are to be expected.

To the limit, it can be said that even if a reactive attitude might be unjustified at the normative level, it is still important to recognize its role in our moral psychology of attachments and affiliations. I would feel confounded if Patrick was not any angry at me for not meeting him for dinner; and he would feel confounded too if I did not show any guilt while telling him that I cannot make it for dinner. My excuse is good and justified, but still we can help being affected by the incident. Given our psychology, feeling those emotions is a natural response; and the lack of those feelings points to emotional distance. We expect from people to respond emotionally to us, according to the quality of our bonding. Part of this mutual responsivity involves addressing and recognizing implicit demands.

At this point, then, our approach also separates from Darwall's. For Darwall, forms of affective bonding are related to what he calls "attitudes of the heart". According to him, attitudes of the heart are part of "that aspect of the human psyche through which we are heartened or disheartened, inspired or deflated, encouraged or discouraged, filled with hope and joy or deflated with despair, emptiness, or sadness" (Darwall, 2017). They help us bonding together because they seek reciprocity, personal attachment and connection. Some examples of these attitudes are love, trust, gratitude or personal hope. For Darwall, attitudes are not part of morality. They are not part of the accountability domain and hence they are not deontic. They do not put claims on us, because they must be freely given. Therefore, they do not have a function in morality. However, as we have argued, these ways of affiliative bonding do license the sort of moral emotions that we have just described, which implicitly address and recognize second-personal deontic claims.

We humans have a need to establish long-term bonds because it has been evolutionarily adaptive (Cheney & Seyfarth, 2008; Seyfarth & Cheney, 2012; Tomasello, 2016). Starting with the affiliative bonding to our parents, which is related to our long dependence on them in development, we come to establish a variety of affective bonds with others along our lives. This bonding involves prosocial preferences for those we are bonded with, but also sensitivity to their demands on us. It is not that we have abstract prosocial preferences for anyone, but rather we develop particular prosocial preferences for those with whom we come to establish affective, long-term relations. Attitudes of the heart contribute to this bonding through the moral emotions they give rise to. These emotions involve implicit forms of normative assessment, in the form of appraisal. And this is why the precursors of both second-personal relations and morality might be found in this kind of pre-normative bonds, promoted by attitudes of the heart.

In sum, those moral emotions which appear in the context of affectionate relationships, work as a way to address justified demands, as Darwall and Strawson notice; but also as a way to interact second-personally with others, and to connect with them. Indeed, Strawson's reactive attitudes presuppose a relationship which is interactive and which involves a second-personal way to relate to others. Accordingly, moral emotions illustrate the second-person standpoint because they ensure the commitment that characterizes intersubjective relationships; hence, they become a bridge to morality. From this point of view, morality emerges out of a group of interrelated individuals; and therefore moral emotions become the intersubjective grounds of morality.

7. The limits of the moral community

The naturalistic turn of Darwall's second-person standpoint of morality can also provide an answer to the question about the limits of the moral community. This answer avoids speciesism –a question that can barely be raised within Darwall's analytical project. Certainly, within the framework of the second-person standpoint of morality, it may be said that the members of the moral community are those who are second-personally competent. Second-personal competence is "the capacity to make demands on oneself from a second-person standpoint: in being able to choose to do something only if it is consistent with demands one (or anyone) would make of anyone (hence that one would make of oneself) from a standpoint we can share as mutually accountable persons"

(Darwall, 2006, p.35). In other words, it is the capacity of the subjects "to determine themselves by these [second-personal] reasons" (Darwall, 2006, p.21), and to enter into relations of mutual accountability with other subjects (Darwall, 2006, p.33). Second-personal competence is what makes us subject to moral obligation, and what gives us an authority to make claims and demands of one another as members of the moral community.

Darwall agrees that acting according to second-personal competence requires some psychological capacity for perspective taking and psychological attribution –a capacity that he envisions in terms of "simulation" or "imaginative projection" (Darwall, 2006, p.44-45). It also requires the ability to assess the situation from an involved other's standpoint, as the precursor to an impartial perspective. Subjects must be able to consider another's standpoint and "compare the responses that one thinks reasonable from that perspective with the other's actual responses, as one perceives them third-personally" (Darwall, 2006, p.48). Besides, subjects must be able to regulate themselves by claims, demands, and norms in a spontaneous way, as all these assumptions and dynamics are taken for granted without necessary awareness of them (Darwall, 2006). Therefore, the question whether a class of individuals count as moral subjects can be reformulated in terms of whether they have the capacity to claim and respond to claims; to recognize another's authority to claim and being recognized the same authority by others; to hold and being held accountable; and to enter in this kind of second-personal, spontaneous interactions.

In this way, Darwall's approach offers a way to proceed in order to determine the limits of the moral community: an individual will count as a moral subject if she shows the relevant psychological features described. The question, then, is whether any non-human animal exhibits the same sort of psychological capacities on which such competence depends. Some of them have already been mentioned: perspective-taking, intentional attribution, emotion expression and emotion recognition. This strategy would parallel the one followed by Gomila (2001b), and Gómez (1998), with respect to whether great apes comply, and to what extent, with the conditions of personhood specified by Dennett (1976). However, the difficulty with this strategy when applied to Darwall's second-personal competence is that the presence of the psychological requirements of such competence may not be sufficient to guarantee the presence of the competence itself. Chimpanzees might have all the necessary psychological processes for second-personal interaction, but still not be second-personally competent, in Darwall's sense, i.e., moral beings.

An alternative strategy is to directly examine the evidence of second-personal competence in non-human animals. If being a member of the moral community consists in addressing demands to one another through implicit emotional reactions, which already entail some form of normativity, then the task is to find out whether any species exhibit this kind of sensitivity to moral claims, through the appropriate emotional reactions. Taking this approach, the research groups led by Kristine Andrews (Andrews, 2009; Vincent, Ring, & Andrews, 2019), Mark Bekoff (Bekoff, 2004; Pierce & Bekoff, 2012) and Frans de Waal (Brosnan & de Waal, 2003; de Waal, 1996, 2006, 2014) have argued that non-human primates do relate through normative expectations, and do react with anger and rage when those normative expectations are transgressed. However, this interpretation is controversial. Michael Tomasello and his group are more skeptical about non-human animals' second-person competence. According to them, we have enough negative evidence enough to deny any normative capacity in non-human animals. The behaviors that the first group take to show resentment are also compatible with disappointment and frustration (Engelmann, Clift, Herrmann, & Tomasello, 2017; Engelmann & Tomasello, 2017; Tomasello, 2016). Regardless of the way the evidence turns out, the point here is that Darwall's notion offers a fruitful path to study the question of the limits of the moral community.

8. The emergence of group norms

Another important question, both in Darwall's analytical project and in our naturalistic interpretation of it, is the relation between the demands addressed within the dyadic structure of second-person relationships, and the explicit norms and codes the social groups of humans develop in time.

In Darwall's project, there is no difference between the valid demands in the dyad, and the universally valid ones; the categorical imperative that justifies the demands, according to their universalizability and impartiality, is already assumed in the intersubjective relation. For the interaction to occur in the first place, participants need to be committed to a categorical and universal principle (de Maagt, 2018). If I engage in a second-personal relation with someone else, I assume that the other will comply with my demands only if they sees them as justified. In Darwall's words, "you and she commonly presuppose that she can freely comply if she finds your request or demand one she could not reasonably reject, regardless of what she desires or how strongly she desires it" (Darwall, 2006, p.245). Yet in everyday practices, an addressee might not recognize the demands of an addresser. For instance, my demand as a customer of being attended in English might not be recognized by the shop-assistant, who might consider that my claim violates his dignity as a French speaker. Or, as Corbí (2005) forcefully describes, a victim might not be recognized by a torturer; i.e. the torturer might deny the victim's experience of the harm that they themselves is causing. Hence, both the shop-assistant and the torturer might see the demands of their customers, and victims respectively unjustified. This lack of agreement about the validity of demands is possible within a group; much more so when the relationship is established between members of groups that disagree about the relevant norms in the first place.

To explain what justifies a demand, Darwall provides an analytical answer, resorting to a version of Smith's impartial spectator (1759). According to Darwall, a demand is justified if any member of the moral community would accept it. Justified blame is addressed as if from an impartial third-party. Consequently, although second-personal demands are described as happening in the dyad, they have already universalizing tendencies (Darwall, 2008); they are expressed as if any member of the moral community could legitimately address it to another member whoever. Therefore, Darwall does not need to explain how the norms accepted within the dyad extend to the other members of the moral community. In Darwall's view, the universality of the demands is prior to the dyad. As already said, the validity of the demand does not depend on its being addressed.

As already pointed out, though, the assumed impartial spectator's view cannot be taken for granted. This strategy faces two sorts of related problems: there is no such an impartial spectator's perspective; and different communities can have different moral codes. First, as we mentioned in section 5, there is no such a generalized or impartial perspective. No individual, as rational as they might be, can adopt such generalized, or impartial, perspective and conclude whether a claim is justified or not. Second, if there were such a generalized perspective, a unique set of demands could be justified. Yet, in practice, what happens is that different communities find legitimate different sets of demands, because they accept different moral codes.

Both issues cannot be solved within Darwall's approach. A naturalist approach, on the contrary, offers a way to raise the question of how reciprocal demands become valid through the group consensus. Allan Gibbard dealt with the question of the emergence of group norms, and the possibility of moral disagreement, from such naturalist standpoint. According to Gibbard (1982, 1989), moral norms emerge out of interaction with others. When we interact, we stick to our positions and avow them; but we also discuss, and look for a consensus. As a consequence of this interaction, "a group will form a community of judgment, and different groups may form different communities of judgment, at odds with each other" (Gibbard, 1989, p.177). In the different communities, there will be discrepancies on local norms; whereas there will be consensus on some restricted topics "on which one needs agreement" (1989, p.179). Accordingly, beliefs about justice will be "beliefs about what the consensus will be on what is just"(Gibbard, 1982, p.42).

The naturalized second-person approach can still resist this view by pointing out that there is no need to foresee a communitarian process of explicit normative debate and consensus. Given that each agent can establish dyadic interactions with many others, the web of dyadic interactions influences which demands are recognized as justified. Thus, it emerges a sort of collective equilibrium. Once this equilibrium is reached, the authority of each subject to address demands based on reasons is warranted by the authority of the group and, finally, by the moral community, who has the role of accepting and approving demands, and requiring respect of members, in cases of disagreement (Corbí, 2005).

From this perspective, the objectivity of moral norms can be compared to the objectivity of the norms of grammar[4]. Grammar norms of a language are objective in the sense that we cannot make them up. Yet they are not mind-independent because they depend on the minds of the speakers of that language. The same holds for moral norms: they are objective in the sense that we as particular individuals cannot change them; but they are not mind-independent because they depend on the dynamics of intentional interaction of the agents. Thus, intersubjective interaction allows us to account for both the presence of group norms, and the disagreement about them among different communities.

9. Conclusion

Darwall's second-person standpoint of morality is undoubtedly a valuable contribution. Its central idea is the attempt to ground moral obligation in mutual recognition between subjects. Yet its analytical and Kantian nature prevents it from developing its potential as an account of our moral psychology. In this paper, we have developed Darwall's proposal from a naturalistic standpoint, and argued that such understanding of its central idea helps explain a variety of features of morality: the motivational power of moral judgments; the special obligations we recognize towards particular subjects; the intersubjectivity involved in moral emotions; the limits of the moral community; and the emergence of moral, group norms

In any case, Darwall helps to make clear that morality is not in the business of strategic self-interest, but the key to the significant interpersonal relationships that make our lives unique and valuable.

[4] We are indebted for this example to Shelly Kagan.

CHAPTER 3. THE ROLE OF LOVE IN OUR MORAL MOTIVATION TOWARDS FRIENDS

"It is the time you have wasted for your rose that
makes your rose so important"

(Antoine de Saint-Exupéry, *The Little Prince*)

1. Introduction

In the middle of her morning news show, the journalist Robin stands up, picks up her stuff and hastily leaves. She has received a call telling her that her friend Ted just had a traffic accident and is at hospital by himself. Watching this scene from the sitcom 'How I Met Your Mother', we the audience assents, consider that as a valid reason, and approve her action. She feels that she cannot but go: it is her friend, she *ought* to go. The fact that a person that she loves is in trouble moves her; and it also counts as a valid reason for her to abandon her work responsibilities. Furthermore, we would consider it morally reproachable if she decided to stay despite her friend's need of help. This is indeed what happens with another character who decides to stay at their appointment meeting. Were Robin to act like this, her friends would feel resentment towards her, and she would probably feel remorse, and guilt.

There are two aspects worth considering in cases like this one: a motivational, and a normative one. Regarding the motivational aspect, Robin leaves because she feels terrifically motivated to help her friend Ted. We are not surprised by her reaction, because we can expect it given the situation, and the close connection between them. Regarding the normative aspect, the fact that Robin's friend is in trouble counts for her as an overriding reason to evade her responsibility to host the news show. We also consider it a valid reason, one that justifies her behavior. The fact that her friend is in trouble both motivates and justifies Robin's leaving the set.

The question that arises from cases like this is: why does Robin have both a motivation and a justification to help her friend in need? The first reply that comes to mind is "because Robin loves her friend". Both her moral motivation and her moral obligation have to do with the particular relationship between her and her friend. Love both moves us to, and provides us with reasons for, action. However, not everybody agrees on this interpretation about moral motivation, and moral obligation. For example, a distinguished view originating in Kant says that although Robin does love her friend, what motivates her to act is her sense of duty, and not her love for Ted; and that what justifies her behavior is that it derives from the moral law, not that it she is acting out of love. Not only them, but especially Kantians insist on this separation between the domain of duty, and the domain of love.

In this paper, we ask ourselves whether the love[5] we feel for our friends[6] plays any role in either our moral motivation to act towards them; or in our moral obligations towards them, that is, in our special duties[7]. Contrary to the Kantian approach, we argue that love plays an essential role in both. To articulate our proposal, we first spell out the Kantian position. In the next section, we present Darwall's second personal version of it. According to Darwall, love does not have a necessary role neither in moral motivation, nor in moral obligation, as both are grounded in the second-personal dynamics of accountability. In section 3, we raise three difficulties for Darwall's proposal: it is psychologically inaccurate, undesirable in practice, and hardly conceivable in theory. To put it another way, we will argue that the Kantian "cold" morality sets, not just a very stringent standard of morality,

[5] We focus on love as an affective state directed at another person. By love we do not mean a kind of universal love to humankind, such as sympathy, compassion, or concern. Claims about the moral function of such a kind of love have already been made (Blum, 2010; Held, 2006; Nagel, 1970; Noddings, 2010; Slote, 1999). Instead, we deal with the kind of interpersonal love that is directed at a person, for the sake of being *that* person (e.g. love for my sister, my friend, or my partner). To put it in classical Greek terms, we do not deal with *agape*, neither with *eros*, but with *philia*; which originally referred to the affectionate regard we have towards friends, family members, business partners, and even a country (Helm, 2013).

[6] For the ease of the discourse, we will only talk about friends. However, the same reasoning applies to siblings, partners, relatives and the like, as far as we love them as we love our closest friends.

[7] In the literature, the duties or obligations that stem from a particular relationship have been called "special duties", or "special obligations" (Jeske, 2014). We indistinctively use "special duties", "special obligations" or even "moral obligations towards friends".

but rather an undesirable one for humans. To solve these problems we propose to depart from the kind of interaction that Darwall notices, and use it to give a second-personal approach to our moral psychology. We contend that both moral motivation, and moral obligation come from our interpersonal relations with particular others, that is, from our second-personal relations with others. From this psychological version of the second-person standpoint, the three problems raised to Darwall's position are clarified, and the role of love is emphasized in both our moral motivation, and our moral obligations towards friends.

2. Moral motivation, moral obligation and love in Kant's picture

In *Groundwork of the Metaphysics of Morals* (1785), in *Critique of Practical Reason* (1788), and in *The Metaphysics of Morals* (1797), Kant does not envision another necessary source of proper moral motivation than the feeling of respect; neither does he envision another source of moral obligation than universalizability. According to Kant, we humans should feel motivated to act morally because of the feeling of respect (5:73, 75); and we are justified to do it, that is, we must do it, if and only if the maxim we are acting upon is universalizable (4:421). In other words, motivation is supposed to flow from justification of obligation.

In more detail, in the Kantian framework, moral motivation is the result of a twofold intellectual process: the awareness of the moral law; and the drive to act according to it. As for the awareness of the moral law, i.e. our moral duty, it takes place through the practical dimension of the pure reason; through what Kant calls "the "fact of reason" (5:31). According to Kant, the mere recognition of the moral law is sufficient to motivate the holy will to act morally (5:32). Yet, even Kant is aware that in non-holy wills the awareness of the moral law might not be enough. As pathologically affected wills, we humans are moved by our subjective desires (4:454). Our natural inclinations do not always follow the verdicts of reason; therefore we need a drive that motivates us to act according to those verdicts. We find this second aspect of moral motivation in the feeling of respect. Due to the moral feeling of respect for the moral law (5:75), we are aware of the greater value of the moral law, in comparison to our happiness, and feel motivated to act accordingly. Thus, both the awareness of the moral law, and the drive to act according to it through the feeling of respect motivate us to act morally, in Kant's account.

Yet Kant was well aware that in the case of humans, the awareness of the moral law and the feeling of respect might not be strong enough to make us humans act according to what we are justified to do. He realized that sometimes we end up doing what is correct because of other motivations, or "moral endowments" (6:399), such as "moral feeling, conscience, love of one's neighbor, and respect for oneself (self-esteem)" (6:399). One of those motivations is love or, as Kant calls it, "mutual love" (6:449), "love of one's neighbor" (6:399) or "benevolence" (5:82).

According to Kant, love has an indirect or secondary role in moral motivation: it is part of those "subjective conditions in human nature that [...] help [people] in fulfilling the laws of a metaphysics of morals" (6:217). It is one of those "moral endowments" (6:399) which make us feel motivated to act morally. However, as love is capricious, changeable, transitory, and biased; it is unreliable as a moral motive (6:470). Therefore, according to Kant, love has a role in moral motivation, but it is just a compensatory one. It can make us do the right thing when the sense of duty, or the feeling of respect, fails to properly motivate us to act morally (shame on us); but it is neither sufficient, nor necessary for moral motivation.

Neither does love ground moral obligation; that I love somebody is not the kind of reason that can make helping her the right thing to do. When acting out of love, I am acting based on partiality, and self-interest; because "love is not anxious about any Inner refusal of the will toward the law" (5:84). Consequently, according to Kant, Robin's justification to help her friend is not that her friend is involved but rather that it is someone who needs her. The maxim of helping a person in need might be a universalizable one which derives from the moral law. Being aware of this justification should move her to help her friend. Only if this motivation fails, love can help so that the right thing is carried out after all.

3. Moral motivation, moral obligation and love in Darwall's picture

In *The Second Person Standpoint (2006),* Stephen Darwall reinterprets the Kantian morality from the second-person standpoint. He grounds both moral motivation and moral obligation not in the individual awareness of the moral law, but in the intersubjective dynamics of accountability. As we will see in this section, according to Darwall, we are both motivated and justified to act morally because we are aware of what others, and we ourselves, can hold us accountable for not doing. The role of love is not necessary in these dynamics, and therefore it is not necessary for moral motivation, neither for moral obligation.

Darwall also follows Kant in his understanding of love as a kind of beneficence, yet he gives an interpretation of it based on the second-person standpoint. According to Darwall (2006), love is a second-personal phenomena. It takes place in interpersonal relationships; it gives reasons to act; it is addressed to persons; it seeks reciprocity; it implies certain duties and expectations; and it presupposes a second-personal relationship between the interactive parties. Love, as an attitude of the heart, is part of "that aspect of the human psyche through which we are heartened or disheartened, inspired or deflated, encouraged or discouraged, filled with hope and joy or deflated with despair, emptiness, or sadness" (Darwall, 2017). It helps us bonding together, and seeks reciprocity. Therefore, both Kant and Darwall see love as an affective attitude that can be directed at any person for the sake of being a person.

3.1. Love and moral motivation in Darwall's picture

Darwall takes the rational individual deliberation that Kant proposes and interprets in terms of the second-person theory. In Kant's account of the fact of reason, we come to recognize the moral law through moral deliberation and, consequently, we recognize our autonomy as lawgiving wills. However, according to Darwall (2009), this "deliberative standpoint alone" (p. 148) only gives us reasons to act according to the moral law, but it does not explain its motivational force: "the most that (first-personal) practical presupposition arguments can show is that a deliberating agent must treat the moral law (and the dignity of persons) as normative reasons for compliance" (p.142). According to Darwall, when these reasons are seen as intersubjective demands, they get an additional authority on us because we feel the responsibility they involve, i.e. we feel accountable or answerable for non-compliance. Hence Darwall's project of the second-person standpoint.

The second-person standpoint is "the perspective you and I take when we make and acknowledge claims on one another's conduct and will" (Darwall, 2006, p.3). The paradigmatic example of such a perspective is the one where a person steps on someone's foot. The person whose foot has been stepped on has a claim right; they can hold the other accountable. The person who stepped on the others' food has the responsibility to comply with the legitimate demand of the other. Both the right to demand, and the responsibility to comply come from the awareness that any member of the moral community would endorse such dynamics. According to Darwall, it is this perspective which is at the core of most of our moral notions. Indeed, moral motivation is defined as "an intrinsic desire to comply with moral demands to which one may be legitimately held accountable, or equivalently, to comply with one's moral obligations" (Dill & Darwall, 2014, p.14). Consequently, he explains moral motivation as a consequence of an implicit understanding of the practice of holding others accountable. The moral subject feels the sense of duty, and it is motivated by it, because he is aware of the possibility of being held accountable, even by himself, if he fails to act morally without excuse (Darwall, 2006). This awareness needs not to be explicit, and it can be implicit in the subjects' emotions, or "reactive attitudes" (Strawson, 1974), such as guilt, remorse or indignation.

Accordingly, what makes the difference for Robin between staying at the set and helping her friend Ted is that she acknowledges what we, and other members of the moral community including herself, would have the authority to demand that she so behave. If she stayed at the set, she would consider our indignation, and Ted's resentment justified. According to Darwall, Robin's moral motivation comes from her second-personal reasoning about what she could justifiably be held accountable for not doing. Hence, moral motivation emerges from accountability, but not from love.

Darwall follows Kant in that love has a complementary role for moral motivation; and that love is neither sufficient nor necessary for moral motivation. Yet Darwall has a more detailed justification for that position based on the difference between respect, and love. According to Darwall (2016), the key difference between the feelings of love and respect is that they are differently related to accountability, and hence they have different roles in moral motivation. Respect is part of the realm of mutual accountability, whereas love is part of the realm of mutual openness. In Darwall's view, love mediates forms of personal attachment and connection that are not essentially deontic; that is, it takes place in relationships where justification does not have an essential role.

For instance, when I resent the person who is skipping in the line, I implicitly demand apologies from that person. By feeling resentment I hold that person accountable, and ask from that person a recognition of my claim; and compliance with their duties towards me as a person. I expect from them to reciprocate the recognition I have shown for them, and that they had violated by skipping in the line. This case is different from the case where I help a friend. When I help a friend in trouble, I do not ask for any attitude in response. I expect that they will reciprocate, but I cannot demand it. As Darwall's emphasizes, love is a freely given attitude, it cannot be claimed; therefore in my helping them I am neither responding to a legitimate demand, nor demanding anything from them in return. As we can see, both kinds of attitudes are similar because they both seek reciprocity. Yet this reciprocity is of a different kind, according to Darwall. In love we expect reciprocity in the good relationship, the attachment and the mutual presence; in respect we expect reciprocity in the form of mutual respect, and in mutual accountability.

As a consequence, Darwall agrees with Kant that love cannot be neither a necessary nor a sufficient source of moral motivation because it is not related to accountability. It only has a complementary role, as Kant says, when the sense of duty comes short in motivating humans because of our imperfect moral psychology. Focusing on the case of our moral motivation towards friends, we might feel motivated to help them as persons qua persons, and hence out of a sense of duty; and also as friends, and hence out of love. Yet, according to Darwall, in both cases the feeling of respect is sufficient to motivate us to act according to our moral obligations; love only helps us to feel motivated when the feeling of duty is not enough, as a sort of deviant cause.

3.2. Love and moral obligation in Darwall's picture

Darwall also denies the role of love in grounding our moral obligations. According to Darwall, a moral obligation is "what we are (morally) responsible for doing, what members of the moral community, including we ourselves, have the authority to demand that we do, by holding us accountable second-personally" (Darwall, 2009, p.149). In other words, it is "what the moral community can demand (and what no one has the right not to do)" (Darwall, 2006, p.20). All the members of the moral community, for the sake of being members of the moral community, can hold another member accountable for incompliance without excuse. Consequently, my moral obligations are what any member of the moral community can hold me accountable for not doing, even if that person is not directly affected by my wrongdoing.

One might say that Darwall's account of moral obligation implies that both the victim and the witness of a moral transgression have the same right to hold the transgressor accountable (Wallace, 2007). Yet Darwall acknowledges that circumstances put people in different positions to claim, and hence under different obligations. If I am in a crowded meeting and I step on someone's foot, this person has a distinctive second-personal authority to ask me to remove my foot, as a person whose foot I stepped onto. Consequently, I have a special obligation towards that person to compensate for the pain I might have caused them. They has an authority upon me which other members of the moral community lack.

Although moral obligations are shaped by the circumstances, as described, Darwall contends that they are still impartial and universalizable. They are what any member of the moral community in that particular situation would have the right to demand; they could be endorsed from a "perspective that we can all share as free (second-personally competent) and rational" (Darwall, 2006, p.276). In the case where I step onto someone's foot, any person in the position of the person whose foot I stepped onto has the right to demand that I remove my foot. Any member of the moral community, including myself, would make that demand, and therefore my moral obligation to comply is universalizable, and impartial. Even more, in Darwall's schema there is no need for anybody, not even myself, to effectively address the demand for the obligation to exist. His project is analytical, not psychological.

In Darwall's picture, then, there is room for special duties, derived from the relationships particular agents get involved in. From this point of view, love might be relevant to our moral duties as long as it may give rise to special duties. Our love might change the circumstances of the people we love, and hence affect our moral obligations towards them. But Darwall does not view love as grounding any duties at all. If we have special obligations towards the people we love, it is not because of love itself; but because our loving them places them in a position which gives them special authority over us. Darwall calls these special obligations "obligations of loving relationship" (2016, p.172), or "duties of relationship" (2016, p.177). They are obligations which are shaped by the loving relationship between the people involved in the dynamics of making and acknowledging claims on one another; and they address the other as "a-person-who-happens-to-stand-in-that-specific-putatively-normative-relation" (2006, p.270).

These special obligations are not actually grounded in love; they are still grounded in accountability. For instance, the justification of Robin's obligation to help Ted is not that she is obliged to help those she loves, but rather it is that any member of the moral community, including herself, could justifiably hold her accountable if she failed to do it without excuse. Furthermore, despite being shaped by love, according to Darwall, these special duties are still universalizable and impartial. They are what any person in that particular relationship should do. Accordingly, the person or moral member of the moral community who occupies the position of being Robin's friend has certain authority for the sake of that position which other members lack. Consequently, what explains Robin's special obligation towards Ted is not love, but the existence of a historical relationship between Robin and him. In this relationship, love changes the moral obligations of both Robin and Ted, yet it does not justify them. Their justification is identical to any other duty.

4. The case of Barney

As we have seen, Darwall contends that love only complements moral motivation, without being necessary; and that it only shapes moral obligations, without justifying them. Against this view, we present the case of Barney from the sitcom 'How I Met Your Mother' as a case study. We slightly change the plot for argumentative purposes so that it resonates with Lawrence Blum's (1980) contraposition between Manny and Dave (p.146-8); and with Kant's example of the man who helps others out of duty because he cannot sympathize with them (4:398). We use this case to raise three concerns to Darwall's view of moral motivation and moral obligation in the context of love and friendship: a psychological problem, because it is psychologically inaccurate; a practical problem, because it is undesirable in practice; and a theoretical problem, because it is hardly conceivable in theory.

Imagine now that instead of calling Robin, Ted calls his friend Barney to ask for help. In the sitcom, Barney frequently treats his social interactions as a game, for which he has laws, rules, and theories. One of Barney's theoretical inventions is what he calls "the Bro code", a set of rules regulating how "bros", good friends, ought to act to one another. And imagine also that, even though Barney loves Ted as much as Robin does, when he acts to help him he does not do it out of love, but out of duty. He spends time with Ted, and he treats him with due care, as Robin does. Yet Barney does not do it "because it is Ted", as Robin does, but "because that is what the Bro code prescribes", or similarly "because that is what friends do". This acknowledgement of what is the moral thing to do towards friends is what motivates Barney to act as a friend. Now, when Ted is in trouble, Barney knows what he is required to do; he recognizes his special obligation towards his friend as a friend and, from this recognition, he feels morally motivated to help Ted. He acts genuinely out of duty; he recognizes his moral obligation and this motivates him to act.

For the sake of the argument, we will assume that the difference between Barney and Robin is that Barney acts purely out of duty, whereas Robin seems to need the feeling of love to comply with what the moral law requires from her. Both Robin and Barney feel morally motivated to help Ted, and both recognize their moral obligation towards him. Yet Barney does not need love to be motivated, neither to recognize his moral obligations. Therefore, in agreement with Darwall's position, it seems that love is not necessary to feel motivated to act morally towards friends, neither to recognize our special duties towards them. Furthermore, if the only reason and motivation that Robin had was love (as we assume), she would help Ted not because that is the right thing to do, as Barney does, but because that is a good thing for her friend, independently of the value of that maxim as a moral law. As moral actions must be motivated and justified by moral considerations, Robin's reason counts as egoistic, and it is hence a reason of the wrong kind to act morally.

It follows from Darwall's view, that Barney's moral psychology is plausible. Although Darwall does not give this example, he defends that love is not necessary to motivate us to help friends in trouble; and that love is not what justifies our moral obligation to do so. In both cases the sense of duty is sufficient. This is a consequence of the required "moral point of view" of the Kantian approach, which "involves abstracting from one's own interests and one's particular attachments to others" (Blum, 1980, p.2). Even if the aim of the Kantian approach is not to describe human relationships, but to assert how these should be; it needs at least to assume that rational and free agents, and a universal and impartial point of view are both possible.

5. The psychological problem: friendship implies motivation for partiality

The characterization of Barney's motivation to help his friend is not an accurate characterization of our moral psychology. Indeed, the Bro code is presented as a gag, but not as something possible in real life. Furthermore, even if it was possible, it would be rare. Our psychology is shaped in such a way that, in friendship, we feel motivated to act out of love, and not only out of duty. To put it in a more clear way, in situations involving a friend, we feel motivated to act because of the particular relationship we have to this other person, that is, out of what we call a "motivation for partiality"; and not just for the norm itself, what we call "motivation for impartiality".

In this context, the pair motivation for partiality and motivation for impartiality resonates with the pair agent-relative and agent-neutral reasons (Nagel, 1986; Parfit, 1984). Agent-neutral reasons are reasons for everyone, as they do not include any essential reference to the person who has them; agent-relative reasons might not be for everyone, as they include an essential reference to the person who has them. Whereas these reasons refer to the justification of an action, the distinction between motivation for partiality and impartiality refers to what motivates it. As Stocker (1976) notices, some moral theories overlooked this distinction, by overlapping both justifications and motivations, and promoted a kind of "moral schizophrenia". Motivation externalists, such as Railton (1984), are aware that what justifies an action might be different from what motivates it. Relying on this distinction, then, we distinguish justification and motivation, and contend that we might sometimes be motivated to act by the binding force of the norm; and sometimes by the concern for the person involved, whatever the justifying reason for that action might be.

In the case of Ted's friends, Robin has a motivation for partiality to help her friend, because she is moved by the fact that the person involved is Ted in particular; whereas Barney has a motivation for impartiality, because he is moved by the implicit norm, explicit norm for him, that friends ought to help each other. Probably when Robin says that she must go "because my friend Ted just had a traffic accident and he is at hospital by himself" the emphasis lies in "my friend Ted" rather than in "friend at hospital". If that is the case, what moves Robin to act is that her friend Ted is involved, whatever the trouble might be. In fact, if someone asked her why she has to leave, they would not be surprised by a reply along the lines of "because it is my friend Ted". This reply is also available to Barney. Yet the difference between him and Robin is that the emphasis in this case lies in "friend at hospital" rather than in "friend Ted". Unlike Robin, what moves Barney to act is the acknowledgement of the norm of what we ought to do for our friends in trouble[8].

It is true that the Kantian position acknowledges that not everyone will act out of duty or, as we put it, out of motivation for impartiality. For instance, we might reject torture out of a motivation for partiality, because we are moved by the pain felt by someone being hurt; or we might reject it out of motivation for impartiality, because we consider it in tension with the moral law. Yet, according to the Kantian approach, even though someone might act motivated by partiality, everyone could, and should, act motivated by impartiality, even in the context of friendship.

In real life this kind of motivation might be found in some institutionalized relationships such as politician-citizen, seller-customer, teacher-student, or doctor-patient. These relationships are based on respect; hence their actions are motivated by the sense of duty, and justified by the principles and norms guiding those relations. Yet in our personal relations to others, as in friendship, we hardly establish the kind of rational, abstract relationship that Kant and Darwall talk about. We can hardly conceive that in real friendship we might be motivated to act only out of sense of duty; neither that we would justify our action appealing to the moral law.

More often than not, we are motivated for partiality to act towards our friends. Friendship involves concern for your friends for their own sake; or, to put it another way, having *de re* attitudes towards them (Alfano, 2017). Besides, as Friedman (1989) notices, the commitment we have with our friends makes us see their interests, ends and values as both justifying and motivating reasons for our actions. Therefore, we end up being motivated to act because of our friends' sake, and we see this acting for their sake as a valid justification for our action.

[8] Notice that the content of Barney's norm might be partial, defending that we ought to treat friends differently. The point is rather that what motivates him is the norm itself, whatever its content might be, and whoever its target.

Furthermore, in emergency cases our spontaneous motivation is for partiality. In the case of a friend in need, we first act spontaneously out of motivation for partiality, and then we think about the morality of our action, if we ever think about it. Our psychology is shaped in such a way that makes us act immediately to help those we love, with no need to consider the permissibility of our actions. Therefore, morality does not start from our capacity to take an impartial point of view, but from our capacity to overlook our self-interest and act in help of those who need us, without further thought.

6. The practical problem: friendship demands motivation for partiality

Not only are we motivated for partiality to act towards our friends, we also tend to think that we *ought* to be. Even conceding, for the sake of the argument, that humans could actually be morally motivated to help friends only out of duty, this is not what we would normatively expect, or demand, from a friend. Indeed, Barney's acting out of the Bro code is presented as an unrealistic element which means to make us laugh. In real life we would be surprised, and even indignant, at Barney's attitude, and Ted would probably resent him. Darwall accepts this as a possible reaction, yet he finds it unjustified. According to him, given our psychology, it is natural that we humans feel sometimes unfitting, or unjustified emotions: such as feeling shame for surviving the Holocaust (Levi, 2017). Therefore, Ted might feel resentment against Barney, but this resentment would not be justified because Barney would have acted morally, for the good reasons.

We morally assess others' emotional responses. In the traditional Kantian view there is no room for moral criticism about emotions, because they are considered to be too changeable, and capricious to ever be appropriate. Yet in our daily life we require appropriate emotional responses from others (Blum, 1980, p.27). This is something that Darwall certainly acknowledges; given his emphasis on the role of the reactive attitudes as a way of implicitly holding others accountable. In other words, according to Darwall, we are justified to blame someone if their emotional expressions are inappropriate, or morally wrong; but we cannot do it if those expressions are simply unfitting.

Leaving apart the ethical debate about whether we are justified in resenting our friends for helping us purely out of duty; it seems that Darwall's account is undesirable in practice. It fails to characterize our mutual normative expectations. We do not want our friends to help us because that is the moral thing to do. This is too detached a picture of friendship. When we ask a friend for help, we want them both to act out of motivation for partiality, and sometimes to do it without further thought.

First, we want out friends to act out of motivation for partiality; we want our friends to help us because it is us. In friendship we want others to act because we are that particular person in relationship to them, and not because our moral obligations have been shaped by some feature of the situation. Even more, we consider that our friends *ought* to act out of motivation for partiality. Therefore, we would feel resentful if a friend told us that he helped us because "that is the right thing to do". In the case of Ted and his friends Robin and Barney, he would probably resent his friends if they visited him at hospital because they would do it for anyone who happened to be in that circumstance, instead of doing it simply because it is *him.*

Second, we want our friends to act without further thought, and we think that they *ought* to. The fact that our friends are in trouble provides us with "one thought too many" (Williams, 1981). We would not like to know that our friends helped us after a long deliberation about the permissibility of the action; neither that they wondered about the permissibility of the action afterwards (Wolf, 2012). As Wolf (2012) puts it, "I don't want my partner to have to think or be concerned about thinking about moral permissibility in order for him to choose to save me" (p.79), her hope being that "he not care so much about the rulings of morality (in these instances) at all" (p.80). When being helped by friends, we want them to be affected by our needs, and be motivated to act accordingly. Only after that we might accept their wondering about the impartial standpoint. Therefore, not only do we want our friends to act because we are the ones in need; but also we want them to do it without further reflection.

7. The theoretical problem: the implicit obligations of friendship

In our view, Darwall's proposal assumes too strong a separation between the affective domain, and the normative one. According to him, our obligations towards our friends bind us because we are aware that we might be held accountable. Our relationship with them is just a sort of contextual feature which might shape the content, and strength of our obligations, but which does not explain their binding force. Thus, according to Darwall, Barney feels bound to comply with his obligations as a friend because he is aware that he could be held accountable, rather than out of a push to help stemming from their friendship relationship.

We find this picture hardly conceivable: our feeling obliged is constitutive of our relationships with friends. It goes with friendship that it gives us obligations, whether there is a moral community who can hold us accountable or not. It is hard to conceive how someone can be emotionally bonded with someone else, but feel morally obliged by the awareness of the possibility of being held accountable, instead of feeling bound by the relationship itself. We feel that we ought to comply with our obligations towards friends because of our friends; not simply because they are "obligations", or "norms". As a matter of fact, there cannot be such a thing as a "bro code" – a set of special obligations between friends –, because special obligations depend upon the particular circumstances of each case and the capacities of the agents involved.

The binding force of our obligations towards our friends is strongly linked with the relationship that grounded its emergence. The special obligations that we have towards our friends find their binding force in the relationship itself, and actually emerge from it. Critically, whereas in the Kantian account friendship only turns our obligations into special obligations, which ultimately derive from the moral law; our claim is rather that those special obligations actually emerge from friendship, and are later on generalized and understood in an abstract way. Friendship is not only a feature which modifies the moral picture, but a whole new picture in which two people grow, bound up with each other, and form normative expectations about each other's behavior. Due to the affective bond that links friends, these normative expectations might be felt as specially binding, and hence experienced as implicit obligations, or norms. Therefore, whereas in the Kantian approach our personal obligations in friendship are derived from impersonal obligations; in our approach our obligations in friendship are indeed personal, and only later on can be experienced as impersonal.

The binding force, and the constitutive dependence on the relationship of our obligations towards friends is revealed specially in cases of failure or transgression. This is exactly what we learn from Barney. In one of the episodes of 'How I Met Your Mother', Barney transgresses one of the rules of his Bro code, and does something wrong to Ted. As expected, he feels guilty: he does not know how to act in front of Ted, he is anxious, and he avoids some conversations with his other friends. He attributes these bad feelings to the fact that he has broken one of the articles of the Bro Code, so he hires his friend Marshall as a lawyer to find a loophole in the code and get him off the hook. What makes the situation funny is that we all know that our moral psychology does not really go like Barney wants, especially in cases where our friends are involved. First, we do not expect our obligations towards friends to be susceptible of being written in a code, as Barney tries. As previously said, our normative expectations towards friends are not like traffic norms, or norms of etiquette, which can be made explicit in a code. They are implicit in our relationships with friends. Second, we all know that what makes Barney feel bad is not that he has broken a rule; was the rule not written in his Bro Code he would still feel bad. What makes him feel bad is that he has not been a good friend to Ted, whatever the Bro code says, and whatever another person in that situation could have done. He will feels guilty because of the wrongdoing he has done to his friend, which has damaged their relationship; not because of his failure to comply with a norm that any person in those circumstances ought to follow, neither because he is aware of the justified accountability demands from the community. Therefore, he can only feel better if Ted forgives him, as his friend Marshall actually tells him:

> Okay, this isn't about the Bro Code, and you know it. The reason that you're upset is because what you did was wrong. And the only way you're ever gonna feel any better about it is if you tell Ted what you did.

What Marshall is actually assuming is that obligations in friendship are of a particular nature: they are constitutive of friendship, and implicit in it.

This implicit nature that we propose to attribute to our obligations towards friends puts them in a blurry area between normative and descriptive expectations. More specifically, special obligations towards friends combine what a friend is expected to do and what they should do, to behave as a friend. They are between the traditional moral judgments, which explicitly apply norms to particular circumstances, prescribing what someone *should* do; and the empirical, descriptive, or statistical expectations, which simply describe what someone *will* do. Despite this fuzzy nature, they are still normative because, as we have seen, their transgression involves a kind of reproach which can be either explicit through verbal speech, or implicit through reactive attitudes. Barney's guilt, or Ted's resentment towards Barney are just some examples showing that a normative expectation is at stake, and hence that in friendship we do assume certain norms and obligations, only that implicitly.

8. An alternative account of special obligations

We have argued that Darwall's account of love as an "attitude of the heart" void of deontic implications does not capture our moral psychology. The difficulties we have raised for his account in the previous section can therefore be avoided when our moral psychology is taken into account. To do so, we propose to adopt a naturalistic second-person approach to morality. As a matter of fact, Darwall's approach, whereas it is analytical and normative, in fact it is committed to a moral psychology. Although this psychological dimension is not his focus, Darwall presupposes certain psychological capacities in the moral agents. However, the kind of moral psychology and of interpersonal relationships that he assumes are abstract and ideal, and it does not capture how our real interactions actually work. In what follows, we will try to provide a more accurate picture of the special obligation that love and friendship ground, taking a naturalistic approach to second-person interactions.

8.1. A naturalistic approach to the second-person

Darwall claims that the rational and free agents that establish a second-personal interaction must be second-personally competent. Darwall briefly describes this competence as including basic perspective-taking, an impartial perspective, some degree of self-control and self-regulation, the ability to hold oneself accountable, and the ability to recognize oneself and others as having shared second-personal authority as mutually accountable second-personally competent agents (Darwall, 2006, 2018).

However, when describing the kind of relationship between these agents, he focuses mostly on the accountability relation, that is, the second-personal interaction. This kind of second-personal interaction is an abstract, ideal one, which does not even need to take place to ground duties. It rather has a justificatory role, and also sets the standard for an interaction to count as really moral. The agents get to feel morally motivated, and to grasp their moral obligations by being aware of what any member of the moral community, including themselves, could legitimately hold them accountable for not doing without excuse. The agents are conceived as partners to a freely agreed contract: their contract requires mutual recognition and the possibility to denounce failures to comply with its agreements. Furthermore, because of the strong separation that Darwall draws between the realm of love, and the realm of accountability, even though this second-personal interaction might take place between two friends, the source of moral motivation and the binding force of moral obligations are still grounded in the accountability relation, not in the affective one. The affective relation only works as a contextual feature which shapes the moral motivation, and the moral obligations of the subjects; the affective relation *per se* cannot ground a moral motivation, or moral obligation, Darwall says.

From a naturalistic second-personal standpoint to morality we delve into the nature of our actual interpersonal relationships. Flesh and blood subjects are not only second-personally competent agents who establish accountability relations. At core, they are social agents who bond with each other, and who establish particular interpersonal relationships to one another. In this context, the phrase second-personal describes a kind of face to face interaction that is mediated by a particular form of mental state attribution, which is mutually and reciprocally contingent, implicit, context-dependent and transparent (Gomila, 2001a, 2002, 2015). This is the kind of interaction, and the kind of subjects, in terms of which we account for our moral psychology when relating with friends.

8.2. Moral motivation, moral obligations, and love

Affective relationships need time to develop. Love emerges out of positive interactions, and gets reinforced through them. These interactions can be of multiple kinds: we might coordinate for a joint enterprise, we might share interest in some event or circumstance, or we might jointly move. By default, intersubjective interaction is rewarding. The result of a trajectory of interactions is an increased, or decreased, affection for each other. This loving affection motivates us to include the interests of our friend as part of our own interests. In this way, we become prosocially motivated. Sometimes, we experience this motivation as deontic: we feel that it is our duty to do what is our hand to help our friends, and expect them to do the same for us. This is what friendship consists in.

In friendship, then, friends might perceive particular circumstances of difficulty or need of the other as a source of a moral obligation to help, and feel motivated to act accordingly because of the relationship itself. Were it not for the affective bond between Ted and Robin, and between Ted and Barney, they would not feel motivated to act in a particular way towards each other, and they would not feel guilty when failing to act according to their recognized duty. It is when Ted and Robin see each other as friends that they mutually adjust to each other, feel motivated to do it, normatively expect it from each other, and react if the other does not comply. It is not that they share a code of mutual obligations; what they share is a commitment to be sensitive to each other's' needs and circumstances, given their capabilities and conditions.

This is why Darwall's view does not match our moral psychology. There is no such a thing as "obligations of loving relationship" (2016, p.172), or "duties of relationship" (2016, p.177). It is the particular circumstances of the situation that might license the acknowledgement of an obligation and the motivation to carry it out. This is especially the case when the protests for incompliance are endorsed by an uninvolved third-party (Isern-Mas & Gomila, 2018). Different friendship relationships may license different duties, and thus those might be differently sanctioned by a third-party. For instance, back to the sitcom "How I Met Your Mother", at least twice in the show Ted feels the duty to provide a shelter to Robin, even if nobody else would go that far in helping her, and even if nobody would blame Ted for not doing so. Furthermore, he might feel the duty to provide a shelter to his friend Robin, after she loses her job or after she breaks up with her partner, because nobody else is in a position to help her, whereas he does not feel so motivated when it is his friend Barney who loses a job or who goes through a break-up, as he can count on several other helpers.

Darwall's view of love as independent of accountability also overlooks the fact that duties of friends are generally not demanded or claimed, but should be recognized anyway. Special duties, for Darwall, are viewed as duties, but moral obligation in friendship works differently. Friends might not hold each other accountable for their actions or omissions. They need not address each other demands. A friend is one that is sensitive to their friend's needs. If one is not so sensitive, friendship goes away.

In other words, love, and affective relationships in general, entail a deontic dimension. This does not mean that there exists a right to be loved or to have friends, but that friendship partly involves experiencing duties towards particulars others, and be sensitive to their claims and demands. Darwall contends that these duties do not go with love itself, but stem from accountability relationships. Yet friendship is not so well structured and defined; it is not possible to provide a closed list of duties of friendship.

On the other hand, given this influence of the kind of relationship in the emergence of moral motivation, and moral obligation, the role of love is essential. Especially in the context of friendship, love is not just a feature that adds to the picture, and slightly changes the motivation to comply, or the kind of obligations which are already present. Rather is it part of the context in which those motivations, and obligations emerge. In other words, it is not that motivation for partiality, and special obligations come from a slightly change in motivation for partiality, and moral obligations, as the Kantian approach contends. In the context of friendship, those are the original forms in which obligation and moral motivation emerge in the first place. Ted's, Robin's and Barney's motivations and obligations towards each other do not derive from abstract principles, and a general motivation to act morally. Only after several second-personal interactions, among them, and among many other persons, and after observing, and endorsing those dynamics in others, they can get a sense of what motivation, and moral obligations one ought to have in the context of friendship. In other words, only from being motivated for partiality, and from following special obligations can they later be motivated for impartiality, and follow some general obligations.

By adopting this naturalistic second-personal standpoint, we avoid the three problems that Darwall's project had to face: the psychological, the practical, and the theoretical ones. First, we can explain why we feel a moral motivation for partiality in the case of friendship: motivation for partiality is the kind of motivation which appears first in a real interaction with a particular person. Only after several interactions of this kind, and observation of, and intervention in others' interactions, we can reach a motivation for impartiality. Motivation for partiality is not an impulse which comes from our imperfect rationality and which we must overcome; rather it is the first kind of motivation that we acquire in friendship. Only after being motivated for partiality we can generalize and be motivated by impartiality.

Second, our account can explain that we want our friends both to act out of motivation for partiality, and sometimes to do it without further thought. Since moral obligations emerge from friendship, it makes sense that we might have some particular demands in that context that we might not have in others. We demand our friends be motivated for partiality because this normative expectation is built upon how we normally act towards our friends. We normally act motivated for partiality, then we come to expect that this is how "real friends" ought to be motivated. Hence we end up seeing this expectation as normative, and further as an implicit obligation in friendship. Whether we are justified or not in claiming this motivation in our friends, our approach helps us understanding why we have such a demand.

Finally, we do not have the theoretical problem because in our proposal the moral obligation, and the moral motivation we have towards our friends develop through interaction with them. When, after several interactions, we become someone's friends, we also come to grasp which moral obligations we have towards them, and come to feel motivated by those obligations. Friendship without moral motivation and moral obligations is not possible.

9. Conclusion

We have proposed a naturalistic approach to the second-person to account for our moral psychology in friendship. Against the Kantian account, we have argued for a constitutive dependence of friendship and moral obligations, and moral motivation. We contend that it is constitutive of friendship that it goes with a motivation for partiality, that is, a motivation to help our friends because they are our friends; the normative expectation of acting out of motivation for partiality; and special obligations which emerge from the relationship itself and which tend to be implicit, but still binding. Therefore, the role of love in our moral motivation and moral obligations towards friends is not just complementary, but essential in giving rise to moral motivation and moral obligation themselves.

CHAPTER 4. A SECOND-PERSONAL APPROACH TO THE EVOLUTION OF MORAL MOTIVATION

"I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important."

(Charles Darwin, *The Descent of Man*)

1. Introduction

The key question in descriptive evolutionary ethics concerns the evolution of human morality, and specifically the evolution of our capacity for normative guidance[9] (FitzPatrick, 2016). As Darwin already pointed out, we must explain how individuals who are in a "struggle for existence" (Darwin, 1859) can also have a "moral sense or conscience" (Darwin, 1871).

In evolutionary accounts of morality, it is essential to explain, not only where moral norms come from, but also why we feel especially compelled to act according to our moral judgments (Björnsson, Eriksson, Strandberg, Olinder, & Björklund, 2014; Rosati, 2016; Roskies, 2003). We do not just judge according to moral standards: we also feel compelled to act according to such judgments. For instance, we feel horrified at the very thought of a taboo transgression, and we feel strongly obliged to help a friend in need. All these behaviors have to do with the phenomenon of moral motivation. Moral motivation is the part of our moral psychology by which we feel somehow motivated to act in accordance

[9] Following Cela-Conde & Ayala (2007), we distinguish between morality as content, meaning the systems or codes of ethical norms; from morality as structure, meaning the capacity for ethics. We deal with morality as a structure, i.e. as a cognitive capacity, and define it as a cognitive capacity for normative guidance (FitzPatrick, 2016; Gibbard, 1989; Joyce, 2006). More specifically, we understand it as a cognitive capacity which is deployed in interpersonal relations, and which is second-personal in essence (Darwall, 2006; Strawson, 1974).

with our moral judgments (Rosati, 2016). Such judgments need not be fully explicit. They may be implicit, for instance, in our moral emotions. Thus, for instance, when I regret what I did, I am implicitly assuming a negative valuation of it. Failure to behave according to our moral motivation typically gives rise to feelings of remorse and guilt.

The difficulty with moral motivation is that it is not a unique phenomenon. As a matter of fact, we may feel inclined to act morally out of two motivational sources: a motivation for partiality, and a motivation for impartiality. Our motivation is for partiality when we feel motivated to act morally *because of* the particular relationship we have to another person; that is, when we would not be so motivated were the recipient of our behavior anyone in general. This kind of motivation has been described in moral philosophy as acting out of care, love or sympathy. On the other hand, our motivation is for impartiality when we feel motivated to act morally *whoever* is the person involved in the situation; that is, when we would still be motivated were the recipient of our behavior anyone else, regardless of our relationship to them. This impartial motivation is what in moral philosophy has been described as acting out of duty, or respect for the law. In both cases, moral motivation drives us to suppress our own interest, only that in the case of motivation for partiality we do it because of the relationship that binds us to the recipient; whereas in the case of motivation for impartiality we do it because of some impartial standards that apply to any human, or relevant subject in general.

An evolutionary account of moral motivation must account for both motivations. However, some theories of the evolution of morality just focus on the evolution of a partial, or prosocial, motivation (for instance, de Waal, 2008; Trivers, 1971), contending that the impartial motivation somehow emerged as a generalization of the former; whereas others just focus on the evolution of a motivation for impartiality, which they take to be a requisite to talk of morality at all, and which implies our experience of the moral norms as objective (for instance, Gibbard, 1982, 1989; Stanford, 2018).

In *A Natural History of Human Morality* (2016), Michael Tomasello offers a genealogy of both motivation for partiality and for impartiality; or, in his own terms, sympathy and fairness, as motives for moral behavior. His strategy is to characterize impartiality as a form of generalized partiality, following the classical strategy of the moral sense school, best exemplified by Adam Smith (1759). Yet, as we will show, his proposal for the evolution of impartiality as a moral motivation risks circularity, as objected by Stephen Darwall (2018). This difficulty can be avoided if one takes into account a second-person standpoint to morality; but not the analytical and normative one proposed by Darwall (2006), but the naturalistic one called for his theory.

In this paper, we propose a unitary evolutionary account of both partiality and impartiality as moral motivations, based on a naturalistic understanding of Darwall's second-person standpoint of morality. In our view, moral obligation and motivation are grounded in the way we interact with others, both because of the affective relationships we develop with particular others, which involve a motivation for partiality, and because of the demands we reciprocally address and recognize each other, which require some impartial validity. In the next section, we introduce and defend the idea that moral motivation can be for partiality and impartiality. In the third section, we present the second-person standpoint of morality, first as Darwall conceives of it, and then as we reinterpret it from a naturalistic approach. In section four, we introduce Tomasello's evolutionary account of morality and Darwall's objections to it. In section five we argue for our way to avoid those objections. Finally, we derive some corollaries which follow from our proposal.

2. Two kinds of moral motivation

To illustrate the difference between the two kinds of moral motivation that we propose, that is, motivation for partiality and motivation for impartiality, consider the following example. In the popular sitcom 'Friends', Judy Geller is motivated to help her son Ross, because she cares for him. But she is not equally so motivated to help Monica, even though she is her daughter. When Judy helps Ross, she acts morally out of a motivation for partiality; she is motivated to help Ross in particular, because of the esteem she feels for him. By contrast, when she gets to help Monica, she acts morally out of a motivation for impartiality; she is motivated to act because it is her duty as a mother to help her daughter.

Of course, nothing prevents the possibility that an agent is moved by both motivations at the same time, as it is the case with a dutiful and loving mother. It is also possible that both motivations correspond to conflicting moral judgments, as in the famous Kant's case of the man who experiences the conflict between helping a friend who is being pursued by a murderer, versus telling the murderer the truth when he asks us whether our friend has taken refuge in our house. A similarly famous example is found in the story of the Good Samaritan: the man helped a complete unknown because he considered it to be his duty to do so, regardless of any relationship whatsoever among the particular agents involved. This sense of duty is what at bottom justifies an objective view of morality (Enoch, 2018; Kant, 1785; Nagel, 1970; Railton, 1986; M. Smith, 1994), as a source of normativity whose standing and validity do not depend on our preferences.

We might also be motivated to act morally through some other, non-moral, causal paths: as a way to improve our reputation, by strategic calculus to achieve a further goal, or even by coercion. Judy Geller might feel motivated to help both Monica and Ross because she might be socially rewarded with praising, and acceptance; or, conversely, to avoid being socially punished, with gossip, criticism and even ostracism. Yet these motives would not count as moral, because they do not stem from a moral judgment, and are not characterized by the characteristic urge of moral motives. They count as prudential and self-interested reasons.

This dichotomy, i.e. motivation for partiality versus for impartiality, resonates with the classical contraposition in moral philosophy between, on the one side, the notions of duty, respect, or fairness; and, on the other side, the notions of care, compassion, sympathy or love. Most moral theories recognize both sorts of motivation, but draw different relationships between them, and give them distinct status. Remarkably, some moral philosophers consider that morality constitutively requires impartiality (Kant, 1785/1996, 1788/1996; Darwall, 2006; Korsgaard, 2010). Notably, Kant (1785/1996) dealt with this dichotomy in comparing the moral worth of actions motivated by duty, and by care; and contended that an action has moral worth if and only if it is motivated by the sense of duty.

> For, in the case of what is to be morally good it is not enough that it conform with the moral law but it must also be done for the sake of the law; without this, that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also often produce actions contrary to the law. (Kant, 4:390)

Whereas this is a normative claim about the moral worth of actions performed out of duty, in comparison to those performed out of care, Kant (1784/1997, 1797/1996) also acknowledged the role of care in moral motivation, from a descriptive, psychological perspective. He recognized both duty and care as "moral endowments" (6:399), which can motivate moral actions. Therefore, he did not refuse care, or "benevolence", for its lack of moral worth; and he viewed both care and duty, or motivation for partiality and impartiality in our terms, as possible motives for moral behavior:

> But since respect for rights is a result of principles, whereas men are deficient in principles, providence has implanted in us another source, namely the instinct of benevolence, whereby we make reparation for what we have unjustly obtained. [*LE*, 27: 415-16]

Since we are interested in the evolution of our moral psychology, we leave aside the normative question about the moral worth of actions. We focus instead on the evolution of care and duty as our moral motives. We assume both motives as part of our moral endowments, as other evolutionary psychologists propose (de Waal, 2006; Tomasello, 2016), and address the question of their evolutionary origins.

2.1. Motivation for partiality

Humans are motivated to act morally towards others with whom they have a particular relationship, or for whom they care. We are not just strategic and selfish agents; we are moved by sympathy (de Waal, 2006; Tomasello, 2016). Sympathy, or other similar forms of affective attachment, such as "sympathetic concern" or "empathy" (de Waal, 2008), and “pity”, “sympathy” or “compassion” (Hume, 1740/1896; Smith, 1759/2006), are proposed dispositions driving prosocial behavior (de Waal, 2008; De Waal & Suchak, 2010; Engelmann & Tomasello, 2017). Specifically, sympathy is defined as an emotional response that consists of feelings of sorrow and concern for someone’s misfortune (Batson, 2009; Darwall, 1998; Hoffman, 2001; Preston & de Waal, 2002; Wispé, 1986). It involves an other-oriented altruistic motivation, because it motivates the sympathizing individual to help the distressed one, and hence it is cognitively more demanding than bare emotional contagion. Consequently, it is proposed as one of the motives for prosocial behavior among great apes (de Waal, 2008; Engelmann & Tomasello, 2017; Jensen, 2016; Tomasello, 2016), yet its presence in other animals is discussed (for a review, see Pérez-Manrique & Gomila, 2017). This feeling is typically triggered by kin and friends.

The role of care, compassion, or sympathy, in the evolution of morality was already emphasized by Charles Darwin (1871). According to him, sympathy is “the all-important emotion”, and has adaptive value: “for those communities, which included the greatest number of the most sympathetic members, would flourish best, and rear the greatest number of offspring” (p.72). Given the vulnerability of humans when they are born, sympathy makes humans to better take care of their offspring, by reacting to their distress. From kin, sympathy can extend towards cooperative partners, and even towards strangers; yet it still works as a reaction to someone else’s distress. In this evolutionary account, moral motivation is grounded in sympathy, and hence it corresponds to what we have called motivation for partiality. According to Darwin, even what we have called “motivation for impartiality” is based on sympathy. Indeed, according to him, even when sympathy gets its higher level of complexity by involving language, social instincts still “give the impulse to act for the good of the community” (Darwin, 1871, p.41). Yet an account of our motivation for impartiality, that is, our motivation to act just because something is the right thing to do, is missing.

This kind of prosocial motivation, based on prosocial preferences and sympathy, by itself, cannot be properly considered as moral. Even Darwin (1871) conceded this point. For instance, he viewed the behavior of a protecting baboon who rescued a young baboon from a group of dogs as not properly moral yet. To count as moral, it should be motivated by a normative judgment, and not just by sympathy and prosociality. Evolutionary accounts of moral sense consider motivation by prosocial preferences or by sympathy as a starting point of the evolution of morality. Importantly, these accounts defend that these impulses do not disappear once moral judgment, and impartial motivation appear; both in our phylogenetic and ontogenetic development. In other words, according to these accounts, pre-moral phenomena such as prosocial preferences and sympathy became what we call moral motivation for partiality at some point, once morality emerged. An evolutionary account should address this transition.

The question about the evolution of the motivation for partiality is typically approached as first, the question about how to go from individuals who seek maximization of their individual reproductive success to individuals who are also concerned for others; or, to put it another way, whether and how prosocial preferences can emerge in a group of strategic, self-interested, cooperators, who seek maximization of individual reproductive success. The question about the evolution of motivation for partiality is secondly approached as how such other-regarding motivation became genuine moral motivation.

2.2. Motivation for impartiality

Humans also have a motivation for impartiality: we are moved to act morally towards others according to some impartial standards of conduct, out of respect for the norm, or a sense of duty. This statement is twofold: first, we humans experience moral norms as objective, as independent from our desires and attitudes (Goodwin & Darley, 2008, 2010, 2012); and, second, this makes us feel especially motivated to act according to them (Rai & Holyoak, 2013; Young & Durwin, 2013). Tomasello recognizes this motivation that comes from the objectivity of moral norms, and the emotions that go with it. He states that we are not moved just by sympathy for the harmed, but also by resentment against disrespect (Engelmann & Tomasello, 2017, 2019; Tomasello, 2016). Therefore, the question about the evolution of the motivation for impartiality also includes the question of the origins of our experience of moral norms as objective or external, as independent of one's tastes and preferences, and therefore, as valid for any moral subject of the relevant community (Stanford, 2018).

Our motivation for impartiality is based on the binding force that norms have for us. This force is revealed when they are violated, when somebody fails to comply with her duty towards somebody else. In case of transgression, the transgressor might feel guilty; the victim might feel resentment; and the spectators might experience indignation (Darwall, 2006; Strawson, 1974). According to Bicchieri's theory of social norms (Bicchieri, 2016), we show this kind of emotional reaction when someone violates a reciprocal expectation; and the fact that we show these very same reactions to a transgression of a norm indicates that moral norms are instantiated as a set of reciprocal expectations. Therefore, motivation for impartiality might presuppose not only the experience of norms as objective, but also the existence of reciprocal expectations about each other behavior. This is why motivation for impartiality can also be understood as a sense of reciprocity (de Waal, 1996), that is, "a set of expectations about the way in which oneself (or others) should be treated and how resources should be divided" (de Waal, 1996, p.95).

However, such expectations do not amount per se to the motivation to be impartial. The fact that there are norms in play is not enough either, for conventional norms are also social but do not bind us as strongly as moral norms do (Turiel, 1983). As Joyce (2006) contends, even if an individual reacts against a violation of an expectation, this reaction does not amount to that individual "thinking of a negative response as *deserved*, or supposing an act to be a *transgression*, or judging a behavior to be *appropriate*, or considering a trait to be *virtuous*, or assessing a division to be *fair*, or believing that an item is *owned*" (p.93). The expectations involved in the motivation for impartiality are not merely empirical, statistical, grounded in regularities. They are also normative (Bicchieri, 2016). They are seen as "ideal standards", and are used to asses others' behaviors; both to punish non-cooperators, and to choose cooperative partners (Tomasello, 2016). These standards allow individuals to compare a behavior in a given situation with the standard of conduct that they would expect in that situation, and react accordingly. Once in place and acknowledged, these ideal standards can motivate individuals to act according to them, and turn into the well-known universalizing tendencies of morality, the impartial point of view (Tomasello, 2016).

Due to its cognitively demanding nature, the presence of a motivation for impartiality, and of the relevant normative expectations, among non-human animals, is contentious: it is accepted by some researchers (Andrews, 2009; Bekoff, 2004; Brosnan, 2006; Brosnan & de Waal, 2003, 2014, de Waal, 1996, 2006, 2014; Pierce & Bekoff, 2012; Vincent et al., 2019); whereas rejected by others (Bräuer, Call, & Tomasello, 2006, 2009; Dubreuil, Gentile, & Visalberghi, 2006; Engelmann et al., 2017; Engelmann & Tomasello, 2017; Roma, Silberberg, Ruggiero, & Suomi, 2006; Sheskin, Ashayeri, Skerry, & Santos, 2014; Silberberg, Crescimbene, Addessi, Anderson, & Visalberghi, 2009; Tomasello, 2016). In any case, the evolutionary challenge is to explain the emergence of this web of normative expectations and their binding force.

In sum, explaining the emergence of morality implies explaining the emergence of individuals with both a motivation for partiality, i.e. some kind of other-regarding motivation; but also a motivation for impartiality, i.e. the motivation to act according to impartial standards. According to Kant, these individuals would have a stronger motivation for partiality, than for impartiality; hence the need to counterbalance the motivation for partiality with the motivation for impartiality.

3. The second-person standpoint

In our view, as in Tomasello's, a second-person approach to morality has the potential to account for the evolution of both motivations, that is, to be partial in favor of those one holds some form of affective binding with, and to be impartial. The approach to morality as a second-personal phenomenon is owed to Stephen Darwall in *The Second-Person Standpoint (2006)*. Yet the aim of his project is not to give an evolutionary account of morality, not even a descriptive one. He does not even attempt to characterize our moral psychology. Instead, he aims at the analytical project of defining some key moral notions such as respect, obligation, right and wrong as involving intrinsically a second-person standpoint, that is, as being grounded in the relationships between subjects. Despite its analytical nature, though, in our view Darwall's proposal can be developed as a naturalistic project, which can shed light on the psychology of morality, and its possible evolution.

3.1. Darwall's second-person standpoint

The second-person standpoint, according to Darwall, is "the perspective you and I take when we make and acknowledge claims on one another's conduct and will" (Darwall, 2006, p.3). When someone steps onto my foot, I assume that I have a second-personal authority as a person to demand the other one to move their foot. I also assume that they, as a person, has the right to demand something on me; and that we both can hold the other and ourselves accountable if any of us does not comply with the other's demand without excuse. According to Darwall, morality presupposes this second-person standpoint. Morality consists in the practices of holding each other accountable and responding to those claims. Accordingly, he understands second-personal morality "as normative requirements that obligate all moral agents", and which "consists in demands with which second-personally competent agents are mutually accountable for complying" (Darwall, 2018, p.809). Second-personal morality presupposes that the participants can acknowledge each other's second personal authority to raise demands; and that they can also hold the other, and themselves, accountable for incompliance without excuse. Thus, "second-personal interactions always have the seeds of universalism in them" (Darwall, 2018, p.811).

As a consequence, moral obligation is defined as "what those to whom we are morally responsible have the authority to demand that we do" (Darwall, 2006, p.14), whereas moral conscience, or moral motivation, is defined as the "intrinsic desire to comply with moral demands to which one may be legitimately held accountable" (Dill & Darwall, 2014, p.14). Therefore, Darwall equals moral motivation with motivation for impartiality, and dismisses motivation for partiality as a source of moral motivation, as it fails the universalizability criterion. According to Darwall, we are motivated to act morally because we are aware of those actions for which we could legitimately be held responsible; and we perceive moral norms as objective because they are actually objective, that is, they are those norms the violation of which would be justifiably blameworthy. Therefore, moral motivation comes from the experience of norms as objective, not from concern for others.

Apart from not giving an answer to our motivation for partiality, the problem with this view is that it is analytical, and normative at core. It is not a description of our moral psychology, but a prescription of how it should be. Darwall's approach holds for rational and free agents whose access to morality is through deliberation (Darwall, 2009). According to Darwall, a norm is moral if in case of incompliance without excuse, a rational and free agent could legitimately be held accountable for any other free and rational agent who is part of the moral community, including himself. Yet this second-personal interaction is not factual but axiomatic: it does not need to take place, and it does not matter whether it takes place. The value of the second-person standpoint is justificatory. It does not explain the emergence of flesh and blood individuals who are motivated by morality, partially and impartially.

3.2. A naturalistic approach to the second-person

Despite its analytical nature, Darwall's second-person perspective actually presupposes a descriptive dimension: it requires agents with a set of psychological capacities for intentional interaction, such as intentional attribution, for instance. This set of psychological capacities is not made explicit by Darwall, yet it is presupposed in his notion of "second-personal competence". Second personal competence is defined as a "capacity to view oneself and another from a second-person standpoint, in which both oneself and the other recognize one another as having the same shared competence and authority to hold themselves accountable to one another" (Darwall, 2018, p.808). In other words, to be second-personally competent, or to be moral, is to have the capacity to be sensitive to the normative claims that any single member of the community can address to me, including myself. This capacity includes basic perspective-taking, an impartial perspective, some degree of self-control and self-regulation, the ability to hold oneself accountable, and the ability to recognize oneself and others as having shared second-personal authority as mutually accountable second-personally competent agents (Darwall, 2006, 2018).

However, Darwall is not much interested in the question of the psychological make-up of the agents endowed with such a competence. Yet he is well aware that they have to be intrinsically social. As a matter of fact, whereas he is not interested in providing an evolutionary account of the emergence of such agents, he concedes that it "is the object of natural selection under the conditions of obligate collaborative foraging" (Darwall, 2018, p.807-8). The problem with his standpoint, though, is that the process of natural selection does not guarantee the emergence of the sort of rational and deliberative agents his view prescribes; for it might just give rise to rather of blood and flesh subjects who might come short of the ideals of the Enlightenment.

A naturalistic approach to the second person, on the contrary, is interested in the explicit characterization of the psychological capacities of the agents that did evolve. In particular, it is interested in the capacity for mutual intentional attribution, as well as emotional expression and recognition, required for face to face intentional and reciprocal interaction in real time, which our moral concepts presuppose (Christensen & Gomila, 2012). This capacity has also been called "the second-person perspective" of psychological attribution (Gomila, 2001a, 2002, 2015). Morality presupposes agents able of this kind of intersubjective interactions (Gomila, 2008; Isern-Mas & Gomila, 2018). Demands are addressed and honored within these kind of interactions. As a matter of fact, an evolutionary account of moral motivation consists in the effort to make explicit the evolution of these psychological capacities as long as they make possible the emergence of the required forms of motivation.

4. Tomasello on the evolution of morality

4.1. Tomasello's proposal

In *A Natural History of Human Morality* (2016), Michael Tomasello proposes an empirically based evolutionary account of the evolution of morality which tries to answer both the question of the emergence of the motivation for partiality and that of the motivation for impartiality, even though he does not use these terms. His proposal relies on two main, related, points: his well-known case for joint agency as the key to human evolution, and his notion of a second person morality as a basic level of normative regulation that emerged from joint agency.

To begin with, Tomasello contends that morality emerged out of a group of individuals who competed to survive. He departs from the common ancestor of chimpanzees and humans, and proposes that a first transition towards morality was the appearance of cooperation. Cooperation appeared, as it is standardly assumed, through reciprocity, understood as a sort of delayed mutualism, when the short term loss of cooperation was compensated by the long-term gain. Conceding this first transition, the challenge is to explain how self-interested cooperators whose motives for cooperation were prudential gave rise to morally motivated agents.

In order to address the challenge, Tomasello introduces a twist to the standard view of evolutionary game theory of agents as self-interested and prudential. For before engaging in reciprocal cooperation these cooperative individuals were not strategic, rational, isolated players who could opt out of the social dilemmas they faced. They were already social. The common ancestors between humans and *Pan* already lived in social groups, and they related to each other, as kin, friends or cooperative partners, in a similar way as living chimpanzees and bonobos currently do. According to Tomasello, we share with other great apes our concern for those with whom we have close ties. This feature was selected through kin selection, to promote the survival of those carrying our genes; and then reinforced through reciprocity and mutualism, to promote the survival of those who help us. Therefore, when early humans encountered the situation where they were better off if they cooperated they were not individualistic, rational, and strategic agents. Rather, they were already affectively interconnected, bounded, agents already with a set of affiliative relationships. For this reason, their motivation to cooperate was not prudential but other-regarding from the start: they already cared for each other, and were able of empathy.

In this cooperative interaction, agents adjusted their behaviors to each other, formed expectations about others' behaviors, and reacted negatively when those expectations were violated. Yet those expectations were still merely empirical, or statistical: they were about how an agent *would* act. Hence their transgression caused frustration or disappointment (Engelmann et al., 2017). To count as fully moral, interactive partners should form as well normative expectations, that is, expectations about how an agent *should* act and what they *deserved*. Such expectations can be said to be in place when their transgression causes the moral emotion of resentment or indignation which, unlike frustration or disappointment, implicitly addresses a claim to the transgressor, asking for recognition of wrongdoing (Darwall, 2006, 2013a; Strawson, 1974).

Tomasello proposes that this sort of normative demands appeared later in the human lineage as a consequence of the change in the ecological conditions of life in the savannah. In adapting to this new environment, the best adaptive strategy was to cooperate through joint intentional activities. Cooperation became necessary to survive, and its more adaptive form was as a joint intentional activity. According to Tomasello, the seeds of morality can be found in this new form of cooperation because it required the appearance —-in our human ancestors— of three new psychological abilities: cognitive processes of joint agency, social-interactive processes of second-personal agency, and self-regulatory processes from joint commitments.

First, joint agency is the kind of agency that both collaborative agents fall into after engaging in joint intentional activity. Through joint agency, agents get to see the other and themselves as part of a cooperative activity which must follow some standards (role ideals) and which could be done by other agents (self-other equivalence). Second, the collaborative partners adopt a new kind of agency: second-personal agency. Through second-personal agency, the agents recognize both the other and themselves as cooperators who must meet some criteria to be able to collaborate in the future. This recognition emerges through the mechanisms of partner choice, and partner control, which promote both evaluation of others, and awareness of others' evaluation of oneself. Finally, what makes this kind of collaborative activity to stand as a kind of morality is joint commitment. The strength of the joint commitment, that is, the agreement to start a collaborative activity, is manifested in the interpersonal feelings of "ought" of the agents, and in their feeling responsible for one another. But it is also manifested in their reactions towards the other's transgressions, expressed as feelings of resentment or protests.

The key for the emergence of morality, then, lies in the development of these psychological capacities that made possible joint intentional activities. Joint intentional activities "can give rise to a "we over-me" psychology that represents the beginning of all things moral" (Engelmann & Tomasello, 2017, p.11). Tomasello calls this kind of morality which emerges from joint activity "second-personal morality". According to Tomasello, a second-personal morality is a dyadic morality of face-to-face interactions between agents collaborating together, and feeling responsible to one another, as a jointly committed 'we'. In fact, the phrase "second-personal morality" is meant to emphasize the scope of this kind of morality, which is reduced to the dyad and, specifically, to the dyad's collaborative activity. To put it in a nutshell, moral norms first appear within a dyadic scope.

It is at this level that proper partial motivation appears, as it departs from a basic level of normative assessment—it is just that it relates to the dyad only. Roughley (2018) argues that there are two more factors which might generate motivation for partiality: the joint identity that the individuals develop in the dyad; and the self-other equivalence that the two individuals come to realize while cooperating. Yet, as we will discuss below, it is difficult to explain how joint identity and self-other equivalence can emerge from joint activity.

Tomasello then proposes a final transition from this second-personal morality to the emergence of objective morality, and hence to the motivation to be impartial. When two partners engage in joint intentional activity, they get to see each other as part of a cooperative activity which must follow some standards, that is, role ideals about how to perform a part of the joint activity. They also understand what Tomasello calls "self-other equivalence"; both partners know that each could perform the other's role and that the other could perform theirs. Once the partners understand both role ideals, and self-other equivalence, they develop an impartial point of view about the agreements initially reached in the dyad.

To go from the norms of the dyad to the norms of the community, Tomasello resorts to a version of Smith's impartial spectator (1759). According to Tomasello (2016, 2018), the impartial perspective is enhanced through two interrelated processes: generalization of role standards; and awareness of third parties' assessment. According to Tomasello, the clue lies in the fact that the dyadic collaboration "occurs between individuals in a larger pool of collaborators in a loosely structured social group" (Tomasello, 2018, p.825). The impartial perspective is acquired through generalization after several interactions with different partners; and through increasing awareness of "how others in the potential pool of collaborators were viewing, or would view, certain kinds of actions within a collaboration" (Tomasello, 2018, p.825). The force of the commitment is thus enhanced by the potential collaborators who witness the collaboration. The role of these bystanders contains the seeds of the "fully moral kind of objectivity and normativity" (Tomasello, 2018, p.825). With this emphasis, Tomasello explains how second-personal morality "had at least some generalizing tendencies –implicit reference to others in the pool of collaborators– that provided the external reference point needed for participants in a collaborative activity to give socially normative forces their due" (Tomasello, 2018, p.826).

4.2. Criticisms

Darwall has raised two circularity objections against Tomasello's evolutionary scenario. First, second-personal morality must already involve universalizing tendencies to be considered morality at all. To put it another way, it makes no conceptual sense to call morality what only concerns a dyad of agents. Second, joint intentional agency already requires impartiality, and therefore it cannot account for its emergence, on pain of regress.

The first objection is that second-personal morality must already include universalizing tendencies to qualify as morality at all. According to Darwall, even if second-personal processes take place only in the dyad, the demands addressed between participants must be of universal application if they are to count as moral at all. Although moral requirements are grounded in second-personal interactions, they are not "restricted to obligations each party has to the other, within the interaction" (Darwall, 2018, p.810). There is no qualitative difference between obligations within the dyad and obligations towards the group. The demands that the participants of the dyad address to each other mean to be valid for any member of the community. To interact "second-personally", individuals need to assume that they would interact in this same way with any member of the moral community, for the sake of being a recognized moral agent. Accordingly, a second-personal demand is not merely "a bare or naked demand" which is only valid in the dyad, but "a putatively legitimate one" (Darwall, 2018, p.808) which is "committed to presuppositions of universal human morality" (Darwall, 2018, p.809). Tomasello's second-personal morality lacks these universalizing trends.

This is acknowledged by Tomasello, who concedes that second-personal morality has "only partially universalizing tendencies" (2018, p.825) and hence cannot be considered "full" morality. This difference could be seen as terminological, with Darwall using morality for norms that apply to everybody and Tomasello distinguishing a sense of morality for norms that apply only for each dyad of interacting agents. But this move does not help with Darwall's second point.

Darwall's second objection is that impartiality is in fact a requirement for joint agency, and therefore, it cannot emerge out of it. According to Darwall (2018), impartiality was, for the participants in the dyad, "something that the mutual intelligibility of their collaboration was presupposing", rather than "something it was creating" (2018, p.812). The reason for this claim has to do with the contractualist view of morals Darwall assumes, according to which the partners to any enterprise must recognize each other as potential partners (Darwall, 2018). For two individuals to jointly act, they must assume the second-personal authority (in Darwall's sense) of anyone capable of entering into this sort of collaborative activity; they must presuppose that both cooperators have an authority to issue claims and demands which is previous and independent of their joint activity. This kind of independent, impartial, authority is revealed by the fact that both cooperators must assume from the start that anybody can be a partner, and that anybody has a right to refuse the invitation to collaborate. In Darwall's words, "it is only by reciprocally recognizing one another's basic independent second-personal authority that you and I can form a committed we" (2018, p.807). Therefore, joint agency cannot be the source of impartiality. In Tomasello's account, participants develop a motivation to be impartial by engaging in joint intentional activity: they develop recognition respect, commitment, and trust towards others, as they interact with them. Yet, according to Darwall, these are actually preconditions for joint agency in the first place.

In sum, Tomasello's proposal fails to provide an account of how standards, or norms of a kind that justifies the term, can appear within two-person collaborations, and then extend to the community. Darwall argues that a norm has to hold for the community right from the start to count as a norm. Similarly, Tomasello proposes that moral norms emerge out of joint intentional activity, but Darwall objects that the recognition of the normative authority of subjects is a condition of possibility of joint agency in the first place. Since Tomasello's second-personal morality consists in the agreements reached by the dyad, it is difficult to explain both how the dyad can be in a joint activity without such standards; and how those agreements could emerge from, and transcend, the dyadic collaboration of agents. Without such impartial standards, no impartial motivation is possible. Therefore, it seems that Darwall is right and that the evolution of fairness and impartiality are not successfully explained in Tomasello's proposal.

5. An alternative account

Evolutionary accounts are typically prey of circularities (Gomila, 1994). The challenge is to explain the emergence of something that cannot be presupposed or assumed in any way in the previous situation. But, at the same time, some degree of continuity is required between the previous and the new stages. In this section, we present our proposal to overcome the difficulties leveled by Darwall against Tomasello's account, also within the framework of the second person standpoint. The basic idea to achieve this goal is to accept Tomasello's major evolutionary transitions, but to provide a different account of the emergence of morality, as implicit normative expectations –which avoids Darwall's objections. Thus, we also start from the precursors of partial motivation in the prosocial preferences that emerge within our ultrasocial strategy, and which became moral once humans became able to judge and apply norms. The second transition took place once such norms were in place, so that something like a sense of duty or respect for the law, which characterizes impartial motivation, could develop.

To articulate this view, avoiding Darwall's objections, two amendments to Tomasello's way to characterize what he calls "second person morality" are required: on the one hand, to realize that the whole community is already involved in any dyadic interaction, as long as any individual can play such dyadic interactions with many others; on the other, that such dyadic interactions may require sanction from a third party, a spectator, any other agent not participating in the interaction. All individuals may play the role of agent in a dyadic interaction and this role as third-party, the positions are interchangeable. Both elements contribute to turning Tomasello's "second-person" morality into a full-blown morality, even according to Darwall standards.

To start with the latter, the key to go from other-regarding preferences to partial moral motivation is to understand the role of the third party in dyadic interactions. Tomasello is actually aware of the role of the third party. As we have seen in section 4.1, Tomasello realizes that the dyadic collaboration is influenced by external viewers who might judge the behavior of the partners, and who might make the partners act in awareness of that possible judgment. We further develop this point by emphasizing that these external agents might not only judge, but critically also sanction, or punish, the behavior of the partners involved in the dyadic interaction. Those involved in the dyad become aware, not only of the possible judgment of the witnesses, but also of their possible intervention, and they might even claim their support. When the members of the dyad are collaborating, they are simultaneously monitoring the perspective of an uninvolved third-party, and checking for their support or lack of support. The two collaborators take the perspective of each other, but also of someone witnessing the interaction. Collaboration might give rise to reciprocal demands for complying whose validity turns on the attitude of the third parties involved. The behavior regularities that might characterize the interaction also depend upon agents external to the dyadic interaction.

Our proposal generalizes over third-party punishment (Fehr & Fischbacher, 2004): that is, when an agent punishes one partner in a dyadic collaboration of which they is not part. In our view, third party sanctions may also take the form of emotional responses. Both sanctioning behaviors and emotional reactions involve an implicit valuation, which provides the basis for the emergence of explicit norms, once language is available.

The difference between Tomasello's proposal and ours is that in his proposal impartiality is supposed to emerge within dyadic interactions, while we view it as the outcome of the group dynamics. For this reason, the standard is properly normative: it is endorsed not just by the collaborative dyad, but also by the community at large. Given this common ground, the collaborative agents can imagine how any other member of the moral community would react to their transgressions, and can reason from an impartial, moral, perspective.

The second element is that individuals establish multiple dyadic interactions, with the multiple members of the group —and establish the role of third party to many others. These multiple and interchangeable roles generate a group level dynamics which typically will reach a normative equilibrium: each dyad, and third party, can count on a common set of expectations about each other behavior. This set of empirical expectations involves also expectations about each other valuations, which get transformed into normative expectations, in the sense that common expectations emerge about how to value some action, when then are shared by the group.

At this stage, partial moral motivation is already possible. Other-regarding preferences become moral motivations once the preference is sanctioned by a shared valuation. Language provides the resources for the full emergence of morality, which is required for the emergence of impartial motivation. First of all, a very basic notion of "good" is enough. Its meaning might be rather indeterminate and polyvalent, but it becomes a tool for valuation. Ontogenetic development provides an indication of what may have happened in the phylogenetic one.

This scenario is inspired by Gibbard's (1989) evolutionary hypothesis of the emergence of normative debate in a primitive society as the way to reach an equilibrium of behavioral preferences. The previous stage of implicit assessment through the reactive attitudes gets transformed by the new possibilities that language offers. Language plays a crucial role in the objectification or externalization process (Gibbard, 1982, 1989; Roughley, 2018; Stanford, 2018). It is through language that we can go from an implicit normativity, such as the one revealed in the emotions that a transgression elicits, to an explicit or objectified one. Once we can put into words those behaviors that we endorse, we can start formulating moral judgments, and hence we can externalize those social expectations shaped by others in interaction. And the motivation to behave for the sake of the norm can take place. As Joyce (2006) notices, Gibbard's picture of the emergence of moral norms in a community is independent of his commitment to expressivism, or to non-cognitivism in general. As a matter of fact, moral judgment may be viewed as a form of cognitive judgment seeking objective validity.

Notice that this account avoids Darwall's two objections. On the one hand, our proposal honors his point that norms involve universalizing tendencies. Norms appear as shared expectations revealed in the valuations implicit in emotional reactions. Empirical expectations become proto-normative when the group dynamics make them stable and independent of any individual in particular. On the other hand, we avoid the second objection, related to the structure of joint agency, as we do not focus on a single form of interaction.

From the picture we have proposed of the evolution of morality, some corollaries follow. First, norms are not mind-independent entities, as moral realists would contend. Norms are the unplanned, unexpected result of individuals' interactions; they are the objectification of the implicit normative expectations that we form about others while interacting with them, when the group acquires its own dynamics. In this sense, they are similar to grammar, the norms of languages[10]: they are not mind-independent, because they actually depend on the minds of the speakers of that language, and can be changed by them; but they are still objective because they cannot be just made up by any speaker alone because they require interaction. Importantly, this does not mean that all norms are "just conventions": due to their intersubjective nature they involve an emotional mechanism of commitment and valuation which makes us feel them as more binding, objective and authority independent (Turiel, 1983).

Second, moral emotions are the bridge from mere interactive and reciprocal adjustment to each other to morality. They express and reveal an implicit level of normativity, and hence they might not require language (Rowlands, 2012a). This kind of proto-normativity might be already present in non-human animals (Andrews, 2009; Bekoff, 2004; Brosnan, 2006; Brosnan & de Waal, 2014; de Waal, 1996, 2014; Pierce & Bekoff, 2012; Vincent et al., 2019), and children (Blake et al., 2015; Blake & McAuliffe, 2011; Castelli, Massaro, Bicchieri, Chavez, & Marchetti, 2014; Engelmann & Tomasello, 2019). Nevertheless, one could say that, although moral emotions are not expressed through language, they actually require propositional content as they are propositional attitudes (Gomila, 2012). Either if they involve an implicit, non-verbal, normativity, or if they involve propositional content anyway; their intermediate position between explicit norms and behavioral adjustments and between subjective preferences and impartial standards maintains.

[10] We are indebted for this example to Shelly Kagan.

Third, moral norms come originally from behavioral dispositions based on emotions, which are lately shaped by interaction with others in a similar way that traffic norms shape our behavior while driving (Sie, 2014). It should not surprise us then to find that we are not always impartially motivated agents, but rather partially motivated ones with some preferences for our "near and dear" (Wolf, 2012); and that we see the moral norms of our group as more objective than those of other groups (Sarkissian, Park, Tien, Wright, & Knobe, 2011). From an evolutionary perspective, morality in general, and moral emotions in particular cannot be that impartial as some expect them to be (Bloom, 2014; Prinz, 2011), and the motivation for partiality must be taken into account. The question about whether an agent with only a motivation for partiality to act morally would count as a fully moral agent becomes a terminological one: with Kantians willing to reserve the term "morality" just for those agents capable of full-blown normative guidance, and those from the school of the moral sense preferring to view it as a graded, fuzzy, term.

Fourth, we can evolutionarily explain why moral judgments are experienced as both motivating, and objective. This apparent contradiction of moral judgments having simultaneously the appearance of both objective statements which state something about the world, and subjective states which motivate us to act, constitutes what Michael Smith (1994) calls "the Moral Problem". Smith's worry in *The Moral Problem* (1994) is to make sense of this paradoxical appearance of moral judgment "with the standard picture of human psychology that we get from Hume" (1994, p.14). Putting apart the discussion about Hume's description of human psychology, our worry here has been of another kind: in explaining impartial motivation, we have provided an evolutionary account for moral judgments being experienced as both objective, and motivational.

Finally, according to this view morality does not emerge to solve "the problem of cooperation" (Greene, 2013). The kind of norms that can emerge from cooperation are just coordination norms, but not necessarily moral ones (Gauthier, 1986). The problem of cooperation can be solved in other ways that do not require morality, such as group selection (Sober & Wilson, 1998), kin selection (Hamilton, 1964), mutualism or reciprocal altruism (Axelrod, 1981; Wilkinson, 1990). To explain why we humans are moral, we need a different starting point: morality emerged not from strategic, self-interested individuals who had to cooperate to survive; but from social individuals who related to each other through second-personal mental state attribution. Accordingly, what was first selected in our species was the need to stablish long-term bonds with others, and to relate to them. The second person perspective of psychological attribution contributes to this, especially in non-verbal creatures. Morality emerged within this second-personal interactions; not because of what joint action required, but because of the set of common expectations developed at the community level, which became normative, as we have presented.

6. Conclusion

We have proposed an account of the appearance of the two motives for moral behavior in the evolution of morality: motivation for partiality and motivation for impartiality. It builds from Tomasello's evolutionary account of motivation for partiality, based on sympathy; and explains the evolution of the motivation for impartiality, which requires the objective experience of norms. Thus, we avoid Darwall's criticisms to Tomasello's proposal.

Our proposal rejects the common assumption that morality emerged in a scenario of strategic, self-interested, cooperators. Instead, morality emerged in cooperators who already had an interest in the wellbeing of others. As we have argued, we have been evolutionary selected to be motivated to bond with others and to take others' interests into account; hence our motivation for partiality to act morally towards others, once we became able of normative guidance.

As Tomasello contends, those cooperators were already tied to others, and motivated by sympathy to act prosocially towards kin, friends and potential partners. And, crucially, they were able of a second-personal mental state attribution. That is, they were able to interact with others through a spontaneous, emotional, and engaged attribution of mental states. Through this sensitivity and adjustment to others, expectations develop. These expectations become normative due to generalization of interactions to the community, and third party endorsement or sanction. After that, these normative expectations become norms because we verbalize them and formulate them as moral judgments. Through this process we came to understand moral norms as objective. Their motivational nature becomes now impartial because it is derived not from the preference for the recipient of the behavior, but from the norm itself.

CHAPTER 5. REPLY TO PRINZ ON THE ROLE OF EMPATHY FOR MORALITY

1. Introduction

The role of empathy within morality has been widely discussed. Some authors consider empathy an essential dimension of morality (e.g., De Waal, 2008; Goldman, 1992; Hoffman, 2001; Masto, 2015; Roskies, 2011), while some others claim that its role has been somehow overvalued (e.g., Bloom, 2014; Maibom, 2009). In this paper we are going to focus on Prinz's Kantian arguments in "Is Empathy Necessary for Morality?" (2011), where he contends that empathy is not a necessary condition for morality because it is not part of the capacities that make up basic moral competence. Presumably he would also reject that empathy is a sufficient condition for morality, as some have claimed, such as Rowlands (2012b) or Masto (2015). But in this paper we will discuss only whether empathy is necessary for morality. Prinz's contributes with a negative answer, and we will defend a positive one.

In the second section, we summarize Prinz's arguments against the necessity of empathy for morality. On the basis of a specific understanding of empathy and moral competence, Prinz concludes that "one can acquire moral values, make moral judgments, and act morally without empathy" (p.213). In the third section, we show that even conceding Prinz his notions of moral competence and empathy, his conclusion does not follow. Prinz's characterization of empathy can still be said to play a role in moral competence as defined by him. Having stated this, we discuss Prinz's understanding of both moral competence and empathy. In section four, we deal with the concept of moral competence, arguing that morality does not reduce to moral judgments. Instead, a morally competent subject is one that feels bound by others' demands in interaction and whose preferences are not only self-interested. Spontaneous affective reactions, such as empathy and moral emotions, are thus also conditions for moral competence. In section five, we further criticize Prinz's notion of empathy because it is oversimplified. We contrast it with alternative notions offered by Batson (2009), Darwall (1998), Masto (2015), Wispé (1986) and De Waal (2008), among others. We also show that Prinz's notion does not apply to some central examples of empathy. Furthermore, we contend that empathy involves a prosocial attitude by highlighting its relation to sympathy. Finally, in the last section, we argue further for a view of morality which takes into account, not just the level of rational judgment and action, but also the level of prosocial preferences and normative interactions. It is at this level that empathy plays its role. Thus, once morality and empathy are properly understood, empathy's role for morality is vindicated.

2. Prinz's argument

The notions of empathy and moral competence are the grounds on which Prinz's argument is built. On the one hand, empathy is explicitly understood by Prinz (2011) as a kind of vicarious emotion: "it is feeling what one takes another to be feeling" (p. 212). Its main requisite is emotional convergence: the vicarious emotion needs to be similar to the one of the perceived subject; i.e. the vicarious and the perceived emotion need to converge. This emotional convergence can occur both automatically or through imagination: we can catch others' emotions either through automatic emotional contagion or through effortful imaginative processes. Consequently both imagination and automaticity are features that might be found in empathy. To talk about empathy, the necessary feature is emotional convergence: you need to feel what it would be like to be in the other's place.

On the other hand, a concrete view of moral competence is implicit across Prinz's paper, but this view is not explicitly explained nor justified. According to him, "empathy is not necessary for the capacities that make up basic moral competence: one can acquire moral values, make moral judgments, and act morally without empathy" (Prinz, 2011, p.213). Therefore, he assumes that moral competence is covered by these three dimensions of morality; otherwise his argument would be that empathy is not necessary for *those* dimensions of morality, and not for morality *period*. Furthermore, what he means by each of these previous terms is very concrete and hanging on moral judgment. Paying attention to his reasoning we can infer that when he talks about moral development he focuses on the acquisition of the ability to make moral judgments; and when he talks about moral motivation he just discusses the motivational strength of moral judgments. Consequently, Prinz's morality is structured around the ability to judge morally: the ability itself, its acquisition in humans and its motivational strength.

Following this conception of morality he argues against the necessity of empathy for moral judgment, for moral development, and for moral conduct. Or, in other words, he argues against the necessity of empathy for the ability to make moral judgments, for its acquisition in human development and for its motivational strength.

First, he contends that empathy is not necessary for moral judgment. According to Prinz, there are cases where "empathy makes no sense" (p.214): the vital organs' case (where after empathizing with five people in need of vital organs, you start questioning whether it is bad to kill an innocent person in order to use their vital organs to save five others), the case of the Rawlsian veil of ignorance (where there is no empathy for the needy, but rather concern for the self), cases where you are the victim of moral transgressions (you do not need empathy to judge the action as wrong), and cases with no salient victim (where you can judge the action without coping with another's suffering). There are other emotions or dispositions, such as disapprobation, that can play a major part in both these challenging cases and the empathy-amenable ones.

Second, Prinz argues that the collected evidence is not sufficient to state that empathy is necessary to develop the capacity to make moral judgments. Against Blair's developmental model, which puts emphasis on empathy, Prinz criticizes the role that Blair's gives to violence inhibition mechanisms (R. J. R. Blair, 1995). According to Prinz, not only are these inhibition mechanisms in a controversial status, but also they cannot account for rules that involve non-violent behavior. In addition, the moral / conventional distinction is supposed to appear in development before the association between empathy and morality. Finally, he claims that the psychopathic condition - which involves moral deficit in these patients - can be explained without appealing to an empathy deficit.

Instead, a more general deficit in moral emotions could explain both the low levels of empathy in psychopaths and the lack of moral competence. Thus, Prinz concludes that Blair fails to establish that empathy is necessary for moral development. Consequently, he prefers to remain skeptical on this point.

Finally, Prinz argues that empathy is not necessary to motivate moral conduct either. First of all, research on empathy in both children and adults is quite weak to assert that empathy leads to action. Secondly, in a moral judgment the motivational impact comes not from empathy but from the emotional basis of the moral judgment itself. It is the emotions that underlie moral judgment which are motivating states. Finally, empathy -defined in terms of vicarious emotion- should have a limited motivational force: the caught emotion is weaker than the originated one, and caught emotions are mostly sadness, misery, and distress, which are not great motivators. When it comes to moral motivation, other emotions -such as those associated with approbation and disapprobation- appear to have a greater impact. Therefore, Prinz concludes that in the case of motivating action from a moral judgment "the meager effects of empathy are greatly overshadowed by other emotions" (p. 220)

Furthermore, and still according to Prinz, not only is empathy unnecessary for morality, it is also pernicious. Empathy has some negative effects: basically, it lacks motivational strength and, it tends to be highly selective because it is influenced by cuteness effects, in-groups biases, proximity effects and salience effects. In addition, it promotes a preferential treatment for those we empathize with, and can be easily manipulated. Thus, Prinz concludes that "empathy has serious shortcomings" (p.227) and that "in the moral domain, we should regard empathy with caution" (p.229).

Whereas the empathic processes that Prinz mentions might not be necessary for his view of moral competence[11], there are reasons to doubt both his characterization of empathy and his view of how a proper morally competent subject is structured.

[11] We discuss Prinz's arguments against the role of empathy in each of the dimensions of morality in section 6.

3. The role of empathy for morality in Prinz's view

Before taking issue with the main argument, we want to begin by relativizing the two main criticisms that Prinz makes of empathy: its high lack of motivational strength and its biases. Regarding the limitations that Prinz highlights, we propose: (1) that whereas the limitations may be true[12], they do not license the inference that empathy is not necessary for morality just because it is not perfect; and (2) that these limitations make sense when the function of empathy is taken into account.

On the one hand, the limitations that Prinz mentions do not prove that empathy is unnecessary for morality. In general, the imperfection of a mechanism does not imply that it is not necessary. For instance, a peacock's tail is not perfect, given that its central function in mating makes it difficult to flight for peacocks. However, it does not follow that the tail is not necessary for flight, even for peacocks, because of its function in keeping stability.

As this example illustrates, natural selection is a satisfacing process that works on what is available and under given constraints. Analogously, it may be the case that empathy is not a perfect mechanism, because of the trade-offs that were to be satisfied in our evolution; maybe it is even a by-product of some other process, a sort of spandrel (Gould & Lewontin, 1979). It does not follow that, because of this imperfection, it is not necessary for morality. Thus, at this first point, we warn Prinz that empathy's limitations only allow us to assume its imperfection, but not its unnecessity.

On the other hand, the noted shortcomings depend on the function that Prinz unfoundedly awards to empathy. In line with his view of morality, Prinz expects that empathy will serve the ability to make objective moral judgments. Prinz assesses empathy according to this implicit criterion and concludes that empathy does not guarantee it. However, at his point Prinz commits the inverse version of the naturalistic fallacy: he infers "is" from "ought". More specifically, from a particular understanding of morality

[12] Find this discussion in section 6.

and empathy, Prinz considers that, to serve morality, a mechanism ought to help making objective and detached moral judgments, and from this “ought” he concludes that empathy is not the mechanism to generate such judgments. As empathy does not fully meet this requirement, Prinz rejects it.

However, if we adopt a different functional standpoint concerning empathy, it may turn out that such limitations are not imperfections. Rather, we must further study these features and verify to which extend they contribute to morality. In fact, the limitations that Prinz finds in empathy suggest that empathy's function is not related to judgment, but to social interaction, an idea that will be developed below.

4. Moral competence

As we have seen, Prinz reduces morality to making moral judgments, acquiring this capacity and being motivated by it. However, there is more in moral competence than judging morally; acquiring the ability to judge morally; and being motivated by moral judgments.

To show this, we can use an adaptation of the thought experiment of Condillac’s statue (Falkenstein, 2010) to think about the conditions for moral competence. Imagine that we teach a robot (or a statue, as in Condillac’s original proposal) some moral principles that he will be able to use to make moral judgments (think of it as a program). Consequently, this robot can reason about moral values, make moral judgments and even act motivated by his moral judgments. But from the point of view of the robot, this is just another program, just some other rules.

As a matter of fact, Prinz is aware that mere judging is not enough for morality. What else should we provide our creature with? If we assume, as Prinz does, that moral judgments have to be intrinsically motivating states, a way to proceed would be to provide an emotional basis to these judgments, because emotions do have motivational power (Gomila & Amengual, 2009). But how is it possible to ground moral judgments in emotions? And how is it that they have this motivational dimension? Notice that emotions are elicited by particular circumstances, not universal properties. Emotions may involve some sort of appraisal of a concrete situation, but this appraisal does not take the form of the application of a universal proposition to a particular case. The valuation may be sensitive to all the particular features of the context. Besides, it is also sensitive to the past history of rewards, and to its value given the current state of the organism and its needs. Thus, those very same emotions whose role in morality Prinz concedes exhibit the same limitations of empathy when a universal standard is assumed. In other words, Prinz needs to justify why partiality excludes empathy from morality, and not emotions in general.

This holds also more clearly for moral emotions - emotions whose appraisal concerns the particular interaction between oneself and the other. In remorse, for instance, one may feel that what one did to another was wrong - but not necessarily wrong in general, but wrong *to* someone (Darwall, 2006). A corollary of this particularism of emotional appraisals is the broad domain they open for moral conflict: one and the same situation may give rise to conflicting judgments, if different reasons are present, but it may also elicit a conflict between judgment and emotional response, or between different moral emotions that can be simultaneously felt. As Masto (2015) points out, most of morally difficult scenarios that we face in ordinary life are not like the generalized ones that Prinz mentions.

What this discussion suggests is that moral judgment is one of the elements of moral competence, but it is not the only one. According to Cela-Conde (1987), morality is made of different levels: the level of motivation, which covers prosocial preferences and second-person mechanisms (Gomila, 2008), such as moral emotions or empathy; the level of normative terms; the level of moral judgment and normative codes; and, the level of ultimate ends and supreme values. From this perspective, Prinz's view is concerned with just one of the levels, and collapses the basic, motivational one, to moral judgment. However, to study the role of empathy for morality we need to investigate which is its role at each level and whether it is part of some of the levels.

Furthermore, Prinz's view of moral competence is simplistic not only from our multi-level perspective. Thus, for instance, Haidt (2008) includes respect for rules and the founding role of the group in his view of morality. Rowlands (2012b), and also Masto (2015) and Monsó (2015), emphasize moral motivation, implicitly and explicitly moral reasons, moral responsibility, and the practice of giving reasons. Finally, Darwall (2006) and Gomila (2008) emphasize the intersubjective dimension of morality: the sense that we are bound to respect other's demands, which we experience as implicitly normative.

A fully morally competent subject, then, should have a sense of normativity, understood as feeling bound by norms implicit in other's demands; and a set of prosocial preferences, which are elicited as spontaneous affective reactions. Being morally competent involves being sensitive to others' needs, and this aspect might require empathy.

5. The notion of empathy

Something similar has to be said as regards empathy, given that there is no agreement on a common definition of it. In fact, as observed by De Vignemont & Singer (2006), "[t]here are probably nearly as many definitions of empathy as people working on the topic" (p.435), which causes "conceptual sloppiness" (Roskies, 2011, 278). The term "empathy" refers to a heterogeneous collection of phenomena (Batson, 2009; Roskies, 2011; Stueber, 2014) with different levels of increasing cognitive complexity (de Waal, 2008). Therefore, our discussion has to involve which characterization of empathy is the best in this context.

The available definitions of empathy in the literature can be distinguished according to two features: (1) whether the consequence of empathy must be emotional convergence, i.e. both participants sharing the same emotion; and (2) whether the causes of empathy must be voluntary and cognitive processes, such as imagination, or automatic and involuntary processes. Prinz's notion of empathy requires emotional convergence, but it is neutral on whether it is caused by voluntary or automatic processes. Yet this definition is not a consensus view.

Both features, emotional convergence and level of automaticity, have been equally criticized and supported along the literature. First, regarding emotional convergence, Monsó (2015) and Masto (2015) see it as a necessary feature of empathy. However, there are some dissenting authors, such as Darwall (1998). According to him, any emotional response which is congruent with another's position should be interpreted as empathy, even if this expressed emotion is different than the perceived one. Secondly, regarding the automaticity of the process, some authors consider that it is a feature that depends on the empathic phenomenon involved (Darwall, 1998); others consider that it is an important feature (Monsó, 2015); and others consider that it is a feature that must not be given, instead the necessary cause of emotional convergence must be imagination (Masto, 2015; Wispé, 1986). Therefore, Prinz should (a) justify his characterization of empathy according to these criteria, and (b) take into account which notion of empathy the defenders of its role for morality assume.

Prinz's notion of empathy applies to some empathic processes but not all of them. Empathy as automatic emotional convergence is also known as "emotional contagion" (Darwall, 1998; de Waal, 2008; Hatfield, Cacioppo, & Rapson, 1993; Hatfield, Rapson, & Le, 2009), "emotional state-matching" (de Waal, 2008), "emotional replication" (Dezecache, Jacob, & Grèzes, 2015), "emotional convergence" (Dezecache, Eskenazi, & Grèzes, 2016), "spread of emotions" (Dezecache et al., 2016), or "lower level empathy" (Stueber, 2014). Empathy as emotional convergence through imagination is related to "projective empathy" (Darwall, 1998), "proto-sympathetic empathy" (Darwall, 1998), "perspective taking" (De Waal, 2008; Decety & Jackson, 2006; Jackson, Meltzoff, & Decety, 2005), or "higher level empathy" (Stueber, 2014), among others. All these processes are not equivalent and, therefore, Prinz's description becomes ambiguous. Consequently, either Prinz's notion requires more concreteness, or its criticism should be put into context for each phenomenon.

Furthermore, there are empathic processes that cannot be tackled by Prinz's notion of empathy, such as "imagine another perspective" (Batson, 2009). And personal distress may count as empathy in Prinz's definition, although it generally does not (Batson, 2009; Maibom, 2009). A position against empathy should also criticize these processes or, at least, justify its removal.

Finally, in its oversimplified characterization of empathy, Prinz tries to distinguish empathy from other phenomena such as sympathy. However, if we consider the history of both terms, this separation turns out to be more complicated.

The term "empathy" ("*Einfühlung*") appeared in philosophical aesthetics to mean the ability to "feel into" works of arts and into nature, namely, expressive perception, so that we project emotional properties on objects that are not able of emotion (Stueber, 2014). From this broad conception of empathy as a human subject's affective participation of an external reality, the concept evolved to address the classical problem of other minds, as an epistemological alternative to Mill's inference from analogy; and to serve the human sciences as the unique methodological alternative to understand subjects and their cultures (Stueber, 2014). Following this view, in psychology, Titchener defined empathy as the subject's awareness in imagination of the emotions of another person (Wispé, 1986). Thus, so understood, empathy is a way to know other minds: a cognitive dimension. As Wispé (1986) calls it, it is "a way of knowing."

As regards sympathy, it was introduced into behavioral sciences by Hume and Smith in discussions of moral motivation and moral development to explain how humans could know, think and feel about the feelings of others (Wispé, 1986). Applied to human psychology, sympathy focuses on human social motivation (Stueber, 2014) and it has been described as "a way of relating" (Wispé, 1986). Specifically, it is defined in moral psychology and moral philosophy as a psychological mechanism which explains how an individual might be concerned about and motivated to act on behalf of another (Stueber, 2014; Wispé, 1986). Thus, sympathy includes two components: cognitive abilities to understand other persons, and emotional and motivational abilities to promote their interests (Stueber, 2014). Thus, in sympathy we find two components: a cognitive one, and an emotional one.

When empathy became a topic of scientific exploration in psychology, both empathy and sympathy merged and empathy absorbed the bidimensionality of sympathy (Stueber, 2014). As empathy had been attributed a role in the recognition and understanding of other subjects (Wispé, 1986), empathy related phenomena were understood as playing an important role in interpersonal understanding and motivating humans to act in a prosocial manner. Empathy's cognitive dimension incorporated sympathy's prosocial character.

Nowadays, "empathic accuracy" means the cognitive phenomenon of apprehension of another's condition, which is related to both "empathy" in its origins and the cognitive component of sympathy; and "emotional empathy" means the emotional reaction to another person who is experiencing or is about to experience an emotion, which used to be the emotionally reactive component of sympathy. As a matter of fact, none of these understandings consider emotional convergence as a necessary feature of empathy, against Prinz's view.

In conclusion, Prinz's concept of empathy is of a very particular nature. It is not the original concept of empathy, which might be equated to empathic accuracy or perspective taking, because Prinz's empathy does not need a cognitive process; neither is it sympathy, which might be equated to compassion, because Prinz's empathy does not require an appraisal, neither a prosocial attitude. Prinz's empathy reduces to the emotional component of empathy, and sympathy; independent of cognitive processes, and prosocial attitudes. It is just an emotional matching.

Given this understanding of empathy, Prinz's claim of the unnecessity of empathy might be reduced to the claim of the unnecessity of emotional contagion; which turns out to be a weaker claim which we would all probably agree with.

However, the common understanding of empathy takes it to involve, not just an emotional reaction congruent to that of the other, but also a prosocial attitude towards the other (Batson, 2009). From this point of view, empathy has motivational force, plus an implicit valuation of the other's situation. Hence, its moral function .

6. Empathy is sometimes necessary for morality

We have shown in section 3 that even conceding Prinz his notions of both empathy and morality, his thesis does not follow. From the fact that empathy has some limitations we cannot infer that empathy is not necessary for morality. Either empathy might not be perfect, which is expectable from the process of evolution through natural selection; or empathy might have a function in morality which is not so straightforwardly related to moral judgment as Prinz assumes. Therefore, we concluded in section 3, Prinz's thesis about the unnecessity of empathy is unjustified.

In this section, we go a step further and argue that not only is the claim about the unnecessity of empathy unjustified, it is also false. Empathy is sometimes necessary for morality, and so are other emotional and interactive phenomena. First, we defend this thesis in Prinz's understanding of both empathy and morality, and show that empathy might prove necessary for moral judgments; secondly, we defend that empathy is necessary for morality in what we consider it to be a preferable framework.

6.1. Empathy is necessary for moral judgment

For the sake of the argument we assume that empathy is a kind of emotional convergence, and that morality reduces to moral judgment. Even in this framework, empathy has a role for morality in the three dimensions that Prinz identifies: in making moral judgments; in learning to judge morally; and in being motivated by moral judgments.

First, empathy is necessary to make moral judgments: it has an epistemological role. As Masto (2015) argues, empathy helps us to manage with nuanced morality; to be more informed; and to know the right thing to do in a given situation. Taking the others' perspective is essential to make moral judgments, since "it does matter, morally speaking, how others actually feel" (p.84). It does not seem so in Prinz's examples because he focuses on paradigmatic cases with already stablished norms. In those cases, such as the vital organs' case or the veil of ignorance, we do not need empathy to know that some action might be wrong. Yet it is in dilemmas, and other cases of conflict, where empathy proves necessary. Evidence for this claim comes from people with autism who describe how they struggle to know how to help someone, despite their motivation to help her (James & Blair, 1996).

Furthermore, in these more nuanced cases it is not only empathy towards the victim what is required, but also towards an impartial spectator. Empathizing with an impartial spectator helps us to grasp what he would judge. In other words, to act morally we need to take into account not only our perspective, but also others' perspectives. As Raitlon (2016) explains, "diminished ability to simulate affectively 'what it is like' for others, or 'what it would be like' for others or for one's own future self were one to take certain actions, leaves one at a systematic disadvantage in successful navigation of the human landscape" (p.7). Social dynamics seem to require this impartial or, in Railton's (2016) words, "non-perspectival" standpoint. Consequently, although we agree with Prinz that there *are* "cases where empathy makes no sense" (2011, p.214), there are also many other cases which do require empathy. Even in those cases where there is not a clear victim, empathy might prove necessary to grasp the right action to do. Morality is not only about approving certain action, but also about being justified. Being justified, or as Masto (2015, p.76) puts it "being morally praiseworthy", requires taking an impartial perspective, and this requires empathy.

Second, empathy is necessary to learn how to judge morally, i.e. to acquire moral values, and to be able to make moral judgments. As we have anticipated in section 4, a creature without empathy could still learn a set of norms, and hence judge morally. Yet she would be clueless when facing a new situation. This is why researchers in moral robots are currently focusing on "empathic" robots, which are able to learn, rather than robots with moral norms (Asada, 2015; Lim & Okuno, 2015; Paiva, Leite, Boukricha, & Wachsmuth, 2017). Focusing in moral development, and in line with what we previously said about the epistemological role of empathy in moral judgments, Railton (2016) states that moral learning might require taking others' perspectives through empathy. To back up his argument, Railton mentions that early damage in regions with a key role in affective simulation and evaluation, such as ventromedial prefrontal cortex (vmPFC) and frontopolar cortext (FPC), can cause serious impairment in moral learning (Baez et al., 2014; Mendez, Anderson, & Shapira, 2005).

Furthermore, moral development does not reduce to learning how to judge morally; it implies learning how to react to certain situations, and also acquiring a sense of normativity. In this sense, empathy has a role. It helps us to connect with others, and interact spontaneously with them. Actually, Railton (2016) mentions that empathy might work as an "alarm signal" (p.7) to call our attention to people who might need help. Indeed, at the sight of someone in need, adults show first a distress-like response and right after that cognitive and affective responses associated with taking others perspectives (Thirioux, Mercier, Blanke, & Berthoz, 2014). Hence, empathy helps us to have a sense of normativity about what we ought to do, which is triggered spontaneously and in interaction with others.

Finally, empathy is necessary for moral motivation. As Heyes (2018) reviews it, empathy in its different forms "motivates helping and consolation behavior" (p.502). According to Prinz, what makes the moral judgment motivating is its emotional basis. Yet how can a moral judgment have an emotional basis without empathy? For moral judgment to be emotionally laden we need a mechanism to connect with others. Empathy, together with other affective phenomena, does this job. Hence not only does empathy help us to know the right action to do, but also it makes us feel the binding force of morality by connecting us to others.

6.2. Empathy is necessary for morality

We have criticized Prinz's arguments against the moral role of empathy assuming his understanding of both empathy and morality. In this section, we show that the main limitations that Prinz sees on empathy, its lack of motivational strength and its biases, are not so beyond Prinz's reductionist frame. As we have discussed in sections 4 and 5, morality goes beyond moral judgment, and empathy goes beyond emotional convergence.

In this new framework, empathy together with other emotional and interactive phenomena proves necessary for morality.

Morality is not an ideal category, consisting in making impartial and objective moral judgments from a detached point of view. Morality is a product of evolution(Tomasello, 2016); it is part of men of flesh and blood; and hence it is far from being ideal, complete, or perfect. From this naturalistic perspective, morality is grounded in intersubjectivity, i.e. in second-personal interactions (Gomila, 2008; Isern-Mas & Gomila, 2018). Both evolutionarily and ontogenetically, morality emerges from interaction (Sie, 2014; Tomasello, 2016; Tomasello & Vaish, 2013): when we interact with one another not only do we learn how we and others are expected to act, but also how we and others should act. For instance, I learn that I should not hit my sister because she cries when I do it, and because I feel indignation when someone does it to me. In the same way, I learn that I should help my friend when she is in need, because she might feel indignation if I do not do it, and I know that she would be legitimated to feel so. Hence, it is through interaction that we learn both the content of our moral norms, and importantly the fact that others matter to us.

In our view, empathy has different functions in morality, apart from those related to moral judgment that we have previously sketched. First, empathy allows us to respond emotionally to others, before we can make any moral judgment. Morality does not reduce to a cold, and detached moral judgment; it also consists on affective, and spontaneous reactions. We react against that person who offended us before we explicitly make the moral judgment about that action. Second, empathy allows us to bond with others because by reacting emotionally towards others, either through a cognitive or an imaginative process, we link with them. Prinz acknowledges this role of empathy in establishing tight social bonds, but he focuses on the dark side of it: bullying of outsiders, and motivation for suicide bombing. These are undeniably bad consequences, but the lack of empathy and hence of bonding would have even worse consequences. If morality emerges from interaction, all those processes and mechanisms which ensure interaction become essential for morality, despite its possible negative by-products. Empathy is one of those. Finally, empathy allows us to learn what others expect from us, and hence how we should act. Even the more basic forms of empathy, such as emotional contagion, have a role : they are necessary for higher-order empathy (Heyes, 2018; Iacoboni, 2009; Meltzoff & Decety, 2003; van Baaren, Decety, Dijksterhuis, van der Leij, & van Leeuwen, 2009). Therefore, empathy turns out to be one of the elements which promote interaction, and hence morality.

One could say that other emotional phenomena do a better job promoting interaction, and that these phenomena do not require empathy. Consequently, even in this new framework, empathy might not be necessary for morality. Prinz could probably endorse such a view; enhancing the role of emotions, and diminishing the role of empathy for morality. Our reply to that criticism is that we can hardly imagine how morality could emerge from a creature with emotions but no empathy at all.

First, it is difficult to imagine how someone could acquire the so-called social or secondary emotions with no empathy. To feel guilty I need to be able to put myself into the crying victim's shoes and acknowledge that I am the one to blame for her sorrow; or to feel indignation I need to be able to put myself in the transgressor's shoes and check that I acted wrong although I knew that it was wrong (Dill & Darwall, 2014).

Second, if for the sake of the argument, we imagine a creature endowed with all our set of emotions but with no empathy, could that creature really interact? Interacting requires not only expressing the emotional state that something might cause, but also grasping the contingency of the one who is interacting with us (Schilbach et al., 2013; Trevarthen, 1977, 1980; Tronick, Als, Adamson, Wise, & Brazelton, 1978). My expressing pride at a strike in my bowling game only counts as interaction if my expression is influenced by the others' presence, as it was the case in the study of Kraut & Johnson (1979). Yet my expression of pride in front of the screen of my laptop after submitting an assignment right before deadline does not count as interaction. Given this difference, could a creature with no empathy be able to interact emotionally with others? Interaction requires recognizing the other as an agent who can react towards our expressions, and we can hardly conceive how this is possible with no empathy.

Finally, if for the sake of the argument we accept that our fictional creature can have emotions, and interact emotionally, could she have some kind of morality? For instance, she might be able to react emotionally towards someone's anger with fear; or towards someone's proud with envy. However, this creature would need to learn all these responses on the basis of the effects that those expressions had previously on her. For instance, she would need to learn that when someone shows anger one should show fear in order to avoid being hit by them; or that one someone shows pride it is because they got something we might desire. This situation puts two problems, at least, to the emergence of morality. First, our creature would not have a clue about which emotions fit better. This is a surmountable worry: our creature would just learn it by trial and error. Yet, and this is the second and more troubling worry, the creature could infer anything other than *unjustifiable* expectations. As Masto (2015) puts it, grounding morality in associative learning has worse consequences than grounding it in empathy, since "it is highly unlikely that we have enough moral knowledge to be secure in the practice of conditioning others to feel outrage, anger, or disgust at all of the actions that we now believe are wrong" (Masto, 2015, p.82). Normativity cannot emerge out of emotional interaction with no empathy. Normativity requires taking the others', or even an impartial point of view, to decide whether the other is justified to do what they did. Without this capacity to grasp the others' standpoint or more cognitively to put ourselves into the other's place, morality is not possible.

In sum, from this perspective the two main limitations that Prinz finds in empathy, its lack of motivational force and its biases, might not be so. As for the lack of motivational force, the emotions that Prinz proposes do not do a better job than empathy. Furthermore, even the basic forms of empathy might encourage moral behavior by making us feel bound by morality. On the other hand, the so-called "biases" of empathy are only so (1) if we assess empathy in its contribution to our capacity to make impartial moral judgments; and (2) if we understand impartiality as Prinz does. First, empathy can have other functions as we have seen in this section. Therefore, it should be not assessed only in its contribution to moral judgment. Second, impartiality in moral judgment is wrongly understood by Prinz as meaning that everybody counts the same. Yet from our naturalistic point of view, we have second-personal duties and obligations which might change depending on the people involved. Even children are sensitive to our different obligations towards strangers, parents and friends (Rhodes & Chalik, 2013). Thus, being impartial is not treating everyone in the same way, but not favoring one's own interest against others.

Furthermore, even if we accept this "partiality" as a problem, Heyes (2018) and Railton (2016) claim that the biases of empathy are not an innate and essential feature of empathy, but a product of a learning process. We have more empathy towards the dear and near because we interact more with them. As a consequence, one of the apparently unsurmountable limitations of empathy might be overcome by interacting with agents from other groups (I. V. Blair, 2002; Dasgupta & Rivera, 2008; Pettigrew, 1998; Pettigrew & Tropp, 2006).

7. Conclusion

It might be said that Prinz is right that empathy is not necessary for morality, but only if one accepts his notions of both empathy and morality, and also the function that he attributes to each of them. The problem is that both notions are highly problematic, as we have tried to show. Moral competence is structured around moral judgment; and empathy, around emotional convergence.

First, we have criticized that Prinz equates imperfection and unnecessity, and that he does not justify the criteria (or function) under which he assesses empathy. The fact that empathy is not perfect for a function does not imply that it is not necessary. Instead, either it may be imperfect or it may help another function. Second, we have criticized Prinz's moral competence because it is too focused on moral judgment. Moral judgment is one of the levels of morality, but it is not the only one. Third, we have criticized the notion of empathy because it is oversimplified, it does not take into account all the empathic processes available in the literature, and it forgets about its relation to sympathy.

A proper view of both morality and empathy suggests, on the contrary, that empathy is required to be a moral agent. If morality is grounded in intersubjectivity (Darwall, 2006; Gomila, 2008); the processes which allow interaction turn out to be a condition for the emergence of morality. Empathy helps at the level of prosocial tendencies and second person mechanisms, and this level is more present than abstract moral judgment in our everyday moral experience. Furthermore, if we consider this interactive aspect of empathy and morality, the pernicious aspects of empathy that Prinz mentions lose their value. Empathy is no more a mechanism to achieve objective, and abstract moral judgments, but a mechanism to favor interaction through flexible, spontaneous, and context-dependent responses. In conclusion, when morality is more than judgment about the rightness or wrongness of certain actions, and empathy is more than emotional convergence; empathy's role for morality is vindicated.

What we have outlined above are just some intuitions. A proper defense or criticism about the role of empathy for morality should: (a) analyze the positive effects of its apparently negative features and vice versa; (b) suggest one (or more than one) function for empathy which raises from its features; and, (c) to study the role that each feature might play in each level of morality. Such a huge task might throw light on the debate and help clarify both the notions of empathy and moral competence.

CHAPTER 6. REPLY TO STANFORD ON THE EVOLUTION OF EXTERNALIZATION OF MORAL JUDGMENT

> "Nobody can or ever will comprehend how the understanding should have a motivating power; it can admittedly judge, but to give this judgment power so that it becomes a motive able to impel the will to performance of an action—to understand this is the philosopher's stone."
>
> (Immanuel Kant, *Lectures on Ethics,* AA 27:1428)

We have two qualms with Stanford's target article: one regarding the way he characterizes the question he raises -the objectification of moral norms-, and one regarding the way he answers it -that objectification of moral norms evolved as a strategy to promote hyper-cooperation and to avoid exploitation. On the first count, we show that objectification is a feature of value judgment in general, whatever the domain. On the second count, we argue that this level of moral norms and judgment cannot be decoupled from the level of motivation and preferences, without which it would lack its motivating strength. We conclude by pointing out the implications of these two points for an evolutionary approach to morality.

First, Stanford seems to take for granted his starting point, that externalization is distinctive of moral judgment. In fact, he does not even consider the possibility that other kinds of value judgments also exhibit this feature. But it is pretty clear that they do as well. Take, for instance, aesthetic judgment: saying that something is beautiful entails that everybody should find it so. Since Kant's third *Kritique* (i.e. Kant, 1790/1987; see also Ginsborg, 2014; Zangwill, 2014), it is undisputed that value judgments aim at universal validity, and cannot be reduced to subjective preferences. Other kinds of judgment are not so objectified. Neither the judgment of the agreeable, which simply claims that one likes something, but not that everyone else ought to like it, nor cognitive judgments, by which we ascribe a property to an object, exhibit this objectivity.

It follows from this that the explanandum of the target article -the externalization of moral judgment- is too narrowly set. An evolutionary account of externalization has to deal with the externalization of value judgments in general, not just of moral judgments. Therefore, the explanans proposed misses the real nature of the phenomenon in question, the fact that humans values, not just moral ones, do not reduce to subjective preferences. An evolutionary account of externalization in terms of the benefits of hypercooperation misses the point of externalization as such. The fact that externalization contributes to morality does not imply that it was actually selected for this effect (Gould & Lewontin, 1979). Sophisticated language, theory of mind, and counterfactual reasoning also contribute indirectly to cooperation in the way that Stanford claims for externalization (transmission of information about others' moral commitments), but they are not considered features of morality. Why should externalization be different?

Second, if for the sake of the argument we accept that Stanford's proposal --that moral norms and values externalization promotes group conformity-- can be extended to all sort of value judgments, the question still remains as to how it is that norm objectification manages to do so. In other words, how is it that people's behavior is sensitive to such normative judgments? It is at this point that a link between the level of preferences and the level of norms is unavoidable. While it is clear that moral judgments do not reduce to prosocial preferences -a point already made by Darwin in *The Descent of Man* (1871)-, it is also important to realize that norms and values are not decoupled from motivations either. Again, we miss an explicit recognition of this fact in the target paper, although it is implicitly assumed near the end, when Stanford notices that moral commitments are intrinsically motivating (and carry a demand for intersubjective agreement). Without this link, recognition of a universal, objective, duty might be behaviorally inert --and fear of group exclusion is not the psychological way through which we experience the call of duty. In other words, the real evolutionary quiz is not the externalization of moral norms, but the fact that we feel the pull of the norm.

From this point of view, Stanford's evolutionary scenario is unsatisfactory, since it is concerned just with the level of norms and values, as detached from the level of motivations. It is as if humans come to formulate judgments, to recognize them as moral in character, and automatically become prone to conform to it and to demand from others a similar conformity. What is missing is an answer as to why we feel obliged to comply, why we feel intrinsically motivated to behave according to the judgment, and expect others to feel the same.

A promising answer to this question, we submit, can be found in Darwall's second-person view of morality (Darwall, 2006). In outline, it would go like this: The objective character of moral norms is geared to subjective motivations because such norms are grounded in the patterns of claims and demands that emerge in intersubjective interaction and that a community comes to sanction. From this point of view, the process that explains the link between norms and preferences would be as follows: (a) I interact with another agent with coordination and reciprocity, and out of evolved prosocial preferences; (b) We explicitly or implicitly make demands on each other; (c) We hold each other accountable in case of failure to comply without excuse; (d) We come up with expectations and norms about how others will or ought to behave; (e) We feel motivated to conform to those norms because we know that we can be hold accountable by others; (f) The group comes to sanction those norms and expects everybody to conform; (g) We end up experiencing these norms as moral (i.e., externalized) and at the same time motivated by them.

Taken together, both points --that externalization is common to all value judgments, and that norms are motivating because of their intersubjective grounding—suggest that an evolutionary account should deal, on the one hand, with norm and value externalization as the real cognitive novelty in the human lineage --a novelty that appears to be general rather than domain-specific-- and on the other hand, with the social dynamics through which intersubjective claims and demands come to be externally sanctioned, and psychologically internalized. From this point of view, externalization is not a selected feature to ensure conformity to moral norms because of their efficiency promoting cooperation; rather, externalization is an outcome of the process through which prosocial preferences become normative because of the demands of mutual accountability that mediate the interactions in a species like ours.

CHAPTER 7. GENERAL DISCUSSION AND CONCLUSIONS

> "Can it be, Theaetetus, that we now, in this casual manner, have found out on this day what many wise men have long been seeking and have grown grey in the search?"
>
> (Plato, *Theatetus*, 202d)

1. The big picture

In this dissertation we have studied moral motivation from a second-personal approach. Our core idea is that morality emerges from our second-personal interactions with others. It is through this kind of emotional, and intentional interaction, that is, second-personal interaction, that we recognize the other as an agent that puts demands on us, and that hence makes us feel motivated to comply. We propose that through second-personal interaction we recognize the particular other as an interactive agent first, and only afterwards as a moral subject. Departing from this idea, we have tackled the psychological phenomenon of moral motivation.

We have assumed that moral motivation is a real phenomenon, and have pointed to the experience of guilt as an indicator of it. Furthermore, this experience also points to the relational, social, nature of our moral psychology. When we experience guilt, we experience it as coming from an internal dialogue with ourselves; as from our holding ourselves accountable. This kind of dynamics, second-personal dynamics, is also what is at the core of other moral emotions, such as remorse, or indignation, and what ultimately explains our experience of morality as both motivating and binding. Pointing to this second-personal nature of morality is one of Darwall's contributions.

According to Darwall, we experience moral judgments as motivating because when we deliberate about what we ought to do, we do it as if from a second-personal standpoint.

Therefore, we feel morally motivated to act because we are aware of what any member of the moral community, including ourselves, would have the right to hold us accountable for not doing in that particular circumstance. Similarly, we experience morality as objective, or external, because somehow it is. Moral obligations are what other members can legitimately hold us accountable for not doing; and hence their objective nature.

However, we do not find Darwall's account complete. We feel motivated to act according to our moral judgments, or according to what others could legitimately hold us accountable for not doing. Yet sometimes we feel motivated to act morally because of a more particular feature of the context: the kind of relationship that we have with the person involved in our potential, moral action. This is what we experience in the context of friendship. When we help our friends, what motivates us to act is the fact that those involved are indeed our friends. The main position in moral philosophy is to reject this kind of motivation as a really moral motivation: to count as moral motivation the motivation to act must come from the awareness that this is the morally right thing to do. Contrary to that position, we argue that morality has other levels beyond that impartiality level.

We have proposed two kinds of motivations, both of them "moral". On the one hand, "motivation for impartiality" is the motivation to comply with moral judgments. It is the moral motivation that Kant, and Darwall stand up for. What moves us in this case is our awareness of what is the moral thing to do; either if we think about it in terms of what the moral law prescribes us, as Kant does, or if we think about it in terms of what the moral community can legitimately hold us accountable for not doing, as Darwall does. It is a motivation which comes from our seeing that moral judgment as impartial, as an obligation. On the other hand, we have proposed a new kind of motivation: "motivation for partiality". This kind of motivation comes from our awareness of who the person involved is. It is not merely prosociality, as we experience it as constraining. When we help our friends we do it because "we ought to", not always because we want to. For instance, I do not want to tell a friend that their partner is cheating on them; but I do it because I ought to. This moral motivation does not come from the moral judgment that "friends ought to be honest to each other", as Barney would have written in his Bro Code, but from the fact that my friend is the one whose partner is cheating on them. The moral judgment is implicit in my action, but the emphasis lies in "my friend" rather than "moral" judgment. I am still repressing my own interest, and I am not necessarily acting to comply with my desires; but out of a constraining force from what I am normatively expected to do. Hence the moral nature of the motivation for partiality.

Once these two kinds of motivations stablished, we have tried to articulate them in a unitary account of the evolution of morality. The need for such a proposal is that (1) motivation for partiality is hardly ever accounted for in evolutionary proposals; and that (2) motivation for impartiality is actually better understood in line with motivation for partiality. In our account of the evolution of morality, we depart from prosocial cooperators who were able to interact second-personally with each other. These cooperators interacted with others, and reacted to others' behaviors. Through these interactions, they started to build expectations about how they would act. These expectations became normative due to generalization of interactions with other members of the community, and due to third party endorsement or sanction. Hence, it is from this scenario that motivation for partiality evolved: the cooperative partners had already normative expectations about others' behaviors but both the motivating and binding force of those normative expectations came to the particular other person involved in the interaction. After that, and once our ancestors acquired more sophisticated linguistic capacities, these normative expectations became norms, as we verbalized them and formulated them as moral judgments. Through this process we came to understand moral norms as objective. From this context motivation for impartiality evolved: the motivational, and binding, nature of those normative expectations was moved now to the norm.

An important point of our proposal is that morality is compound not only by moral judgments and motivation for impartiality, as traditionally assumed. Morality has other levels which are indeed necessary even for the impartial level. Prosocial tendencies, moral emotions, empathy, and the like are also part of our moral psychology, and we have the risk of missing their role if we just focus on moral judgment, and moral norms. Similarly, motivation for partiality is not simply a stage of our evolution of morality, or a kind of proto-morality; it is part of our moral psychology and it explains our moral behavior in many situations of our daily life. Focusing on moral judgment makes us lose the big picture of how our moral psychology works.

2. Shortcomings and further lines of research

This dissertation has some shortcomings. Yet, as these shortcomings can be seen as ideas for future work, we present the shortcomings together with these further lines of research.

First, we could have gone in greater depth in the study of the second-personal standpoint. We could have provided a more detailed account about its evolution, development and etiology. As for the evolutionary account, Tomasello has already introduced in the field the concept of "second-personal morality"; we expect that in the coming years more research will follow this trend. In developmental psychology, and in cognitive psychology, there is also a program of research which emphasizes the role of interaction, and intersubjectivity, in our social cognition. It is led by psychologists and philosophers such as Di Paolo, Gallagher, Hutto, Reddy, Schilbach, or Zahavi. As a future project, it would be worth reading their research and integrate it in our project of describing our moral psychology.

Second, we could have looked backwards to the historical grounds of the second-person standpoint. Although we have focused on Darwall's proposal of the second-person standpoint, this proposal draws from several sources. Kant, Fichte, Hegel, Lévinas or Ricoeur are just some of the main authors whose work should be discussed in a project about the philosophical background of the second-person standpoint.

Third, our proposal aims to fairly describe our moral psychology; hence everything we have proposed in this thesis should be empirically tested. Many research questions follow: How can we test whether our moral motivation comes from our relations with others? Do people actually act out of motivation for partiality? Do people share our intuition that in some contexts one kind of moral motivation is required? Could motivation for partiality be found in non-verbal, or pre-verbal creatures? How do motivation for partiality and motivation for impartiality develop in our ontogeny? Do children share our intuitions about what to expect from friends? When do they start acquiring something like motivation for impartiality? Do other species show anything similar to these two kinds of motivations? As we propose a whole account of our moral psychology, and especially of our moral motivation, most of the claims in this dissertation are susceptible of empirical research.

Finally, moral motivation is a core topic in meta-ethics, and moral psychology. In this dissertation we could have gone into more details to the debates in meta-ethics on moral motivation, which divide internalists and externalists, on the one hand; and Humeans and Anti-humeans, on the other hand. Similarly, we could also have related our account of our moral psychology to other accounts such as Greene's, Haidt's, or Rozin's, for some.

EPILOGUE

Academic work has the risk of being too obtuse and abstract. Many times, people out of academia do a better job of accounting for the apparently complex, and abstract phenomena that we try to account for in academia by doing extensive readings, writing long papers, and attending world-wide conferences.

Talking about her experience volunteering at the Vasilika Camp (Greece), Clara summarized, without noticing, the thrust of this dissertation in two passionate sentences whose simplicity still gives me goose bumps. Asked why she volunteered in that camp, she replied:

> Because in a few years' time, this will get into the history books - because it will get into all the history books - and future generations will hold us accountable. And what will we tell them?

As Clara grasped perfectly, morality has do to with being accountable to particular others, not to abstract principles that we become aware of through the exercise of our rationality. This is the message that "the second-personal dimension of our moral psychology" intends to convey through so many words and pages: that our moral obligations, and our moral motivation come from our personal, and particular relationship with particular (second-) persons.

REFERENCES

Alfano, M. R. (2017). Friendship and the Structure of Trust. In A. Masala & J. Webber (Eds.), *From Personality to Virtue* (pp. 186–206). Oxford: Oxford University Press. http://doi.org/10.1093/acprof

Andrews, K. (2009). Understanding norms without a theory of mind. *Inquiry*, *52*(5), 433–448. http://doi.org/10.1080/00201740903302584

Asada, M. (2015). Towards Artificial Empathy: How Can Artificial Empathy Follow the Developmental Pathway of Natural Empathy? *International Journal of Social Robotics*, *7*(1), 19–33. http://doi.org/10.1007/s12369-014-0253-z

Axelrod, R. (1981). The Emergence of Cooperation among Egoists. *The American Political Science Review*, *75*(2), 306–318. http://doi.org/10.2307/1961366

Baez, S., Manes, F., Huepe, D., Torralva, T., Fiorentino, N., Richter, F., … Ibanez, A. (2014). Primary empathy deficits in frontotemporal dementia. *Frontiers in Aging Neuroscience*, *6*(OCT), 1–11. http://doi.org/10.3389/fnagi.2014.00262

Bagnoli, C. (2006). Moral emotions and the vocabulary of mutual recognition. *CxC - Calls for Comments – Sito Web Italiano per La Filosofia*.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414. http://doi.org/10.1098/rsbl.2006.0509

Batson, C. D. (2008). Moral masquerades: Experimental exploration of the nature of moral motivation. *Phenomenology and the Cognitive Sciences*, *7*(1), 51–66. http://doi.org/10.1007/s11097-007-9058-y

Batson, C. D. (2009). These things called empathy: eight related but distinct phenomena. In

J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–16). MIT Press.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*(2), 290–302. http://doi.org/10.1037/0022-3514.40.2.290

Bekoff, M. (2004). Wild justice and fair play: Cooperation, forgiveness, and morality in animals. *Biology and Philosophy*, *19*(4), 489–520. http://doi.org/10.1007/sBIPH-004-0539-x

Bicchieri, C. (2016). *Norms in the wild: how to diagnose, measure, and change social norms*. New York: Oxford University Press.

Björnsson, G., Eriksson, J., Strandberg, C., Olinder, R. F., & Björklund, F. (2014). Motivational internalism and folk intuitions. *Philosophical Psychology*, *28*(5), 715–734. http://doi.org/10.1080/09515089.2014.894431

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. http://doi.org/10.1016/0010-0277(95)00676-P

Blair, I. V. (2002). The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, *6*, 242–261. http://doi.org/10.1207/S15327957PSPR0603

Blake, P. R., & McAuliffe, K. (2011). "I had so much it didn't seem fair" Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224. http://doi.org/10.1016/j.cognition.2011.04.006

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., … Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258–261. http://doi.org/10.1038/nature15703

Bloom, P. (2014). Against Empathy. *Boston Review*, 1–11.

Blum, L. A. (1980). *Friendship, Altruism, and Morality*. (T. Honderich, Ed.)*The Philosophical Quarterly* (Vol. 32). New York: Routledge. http://doi.org/10.2307/2960086

Blum, L. A. (2010). *Friendship, Altruism and Morality*. New York: Routledge.

Bräuer, J., Call, J., & Tomasello, M. (2006). Are apes really inequity averse? *Proceedings of the Royal Society B: Biological Sciences*, *273*(1605), 3123–3128. http://doi.org/10.1098/rspb.2006.3693

Bräuer, J., Call, J., & Tomasello, M. (2009). Are apes inequity averse? New data on the token-exchange paradigm. *American Journal of Primatology*, *71*(2), 175–181. http://doi.org/10.1002/ajp.20639

Brosnan, S. F. (2006). Nonhuman species' reactions to inequity and their implications for fairness. *Social Justice Research*, *19*(2), 153–185. http://doi.org/10.1007/s11211-006-0002-z

Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, *425*(18 september 2003), 297–299. http://doi.org/10.1038/nature01987.1.

Brosnan, S. F., & de Waal, F. B. M. (2014). Evolution of responses to (un)fairness. *Science*, *346*(6207), 1–7. http://doi.org/10.1126/science.1251776

Castelli, I., Massaro, D., Bicchieri, C., Chavez, A., & Marchetti, A. (2014). Fairness norms and theory of mind in an ultimatum game: Judgments, offers, and decisions in school-aged children. *PLoS ONE*, *9*(8). http://doi.org/10.1371/journal.pone.0105024

Cela-Conde, C. J. (1987). *On genes, gods and tyrants: on the biological causation of morality*. Springer Science & Business Media.

Cela-Conde, C. J., & Ayala, F. J. (2007). *Human Evolution. Trails from the Past*. New York: Oxford University Press.

Cheney, D. L., & Seyfarth, R. M. (2008). The evolution of a cooperative social mind. In T. K. Shackelford & J. Vonk (Eds.), *The Oxford handbook of comparative evolutionary psychology* (Vol. 401, pp. 507–528). New York: Oxford University Press.

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1249–1264. http://doi.org/10.1016/j.neubiorev.2012.02.008

Coke, J. S., Batson, C. D., & McDavis, K. (1978). Empathic Mediation of Helping: A Two-Stage Model. *Journal of Personality and Social Psychology*, *36*(7), 752–766. http://doi.org/10.1037/0022-3514.36.7.752

Corbí, J. E. (2003). *Un lugar para la moral*. (V. Bozal, Ed.). Madrid: La balsa de la medusa.

Corbí, J. E. (2005). Emociones morales en la flecha del tiempo: un esquema de la experiencia del daño. *Azafea*, *7*, 47–64.

Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*(1), 100–108. http://doi.org/10.2174/1872210510666161027101910

Darley, J. M., & Latane, B. (1968). Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology*, *8*(4), 377–383. http://doi.org/10.1037/h0025589

Darwall, S. (1998). Empathy, sympathy, care. *Philosophical Studies*, *89*(July 1997), 261–282. http://doi.org/10.1023/A:1004289113917

Darwall, S. (2006). *The Second-Person Standpoint: Morality, respect, and accountability*. Cambridge: Harvard University Press.

Darwall, S. (2009). Why Kant Needs the Second-Person Standpoint. In T. E. Hill (Ed.), *The Blackwell Guide to Kant's Ethics* (pp. 138–158). Blackwell Publishing Ltd. http://doi.org/10.1002/9781444308488.ch6

Darwall, S. (2013a). *Honor, History, and Relationship: Essays in Second-personal Ethics II*. Oxford: Oxford University Press.

Darwall, S. (2013b). *Morality, Authority, and Law: Essays in Second-personal Ethics I*. Oxford: Oxford University Press.

Darwall, S. (2016). Love's Second Personal Character: Holding, Beholding, and Upholding. In E. Kroeker & K. Schaubroeck (Eds.), *Love, Reason, and Morality* (pp. 93–109). New York: Routledge.

Darwall, S. (2017). Trust as a Second Personal Attitude (of the Heart). In P. Faulkner & T. Simpson (Eds.), *The Philosophy of Trust* (pp. 35–50). Oxford: Oxford University Press.

Darwall, S. (2018). "Second-personal morality" and morality. *Philosophical Psychology*, *31*(5), 804–816. http://doi.org/10.1080/09515089.2018.1486603

Darwin, C. (1859). *On the Origin of Species*. London: John Murray.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. (A. J. Desmond & J. R. Moore, Eds.)*Penguin classics*. London: Penguin.

Dasgupta, N., & Rivera, L. M. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Social Cognition*, *26*(1), 112–123. http://doi.org/10.1521/soco.2008.26.1.112

de Jaegher, H. De, & di Paolo, E. Di. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*, 485–507.

de Maagt, S. (2018). It only takes two to tango: against grounding morality in interaction. *Philosophical Studies*, 1–17. http://doi.org/10.1007/s11098-018-1150-3

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10*, 435–441. http://doi.org/10.1016/j.tics.2006.08.008

de Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge: Harvard University Press.

de Waal, F. B. M. (2006). *Primates and philosophers: how morality evolved*. Princeton: Princeton University Press.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu.Rev.Psychol.*, *59*, 279–300. http://doi.org/10.1146/annurev.psych.59.103006.093625

de Waal, F. B. M. (2014). Natural normativity: The "is" and "ought" of animal behavior. *Behaviour*, *151*(2–3), 185–204. http://doi.org/10.1163/1568539x-00003146

De Waal, F. B. M., & Suchak, M. (2010). Prosocial primates: Selfish and unselfish motivations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2711–2722. http://doi.org/10.1098/rstb.2010.0119

Decety, J., & Jackson, P. L. (2006). A Social-Neuroscience Perspective on Empathy. *Current Directions in Psychological Science*, *15*(2), 54–58.

Dennett, D. (1976). Conditions of Personhood. In A. Rorty (Ed.), *The identities of persons* (pp. 175–196). Berkeley: University of California Press. http://doi.org/10.1007/978-1-4612-3950-5_7

Dezecache, G., Eskenazi, T., & Grèzes, J. (2016). Emotional convergence: a case of contagion? In S. S. Obhi & E. S. Cross (Eds.), *Shared Representations: Sensorimotor Foundations of Social Life* (pp. 417–438). Cambridge University Press.

Dezecache, G., Jacob, P., & Grèzes, J. (2015). Emotional contagion: its scope and limits. *Trends in Cognitive Sciences*, *19*(6), 297–299. http://doi.org/10.1016/j.tics.2015.03.011

Dill, B., & Darwall, S. (2014). Moral Psychology as Accountability. In J. D'Arms & D. Jacobson (Eds.), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics* (pp. 40–83). Oxford University Press.

Dubreuil, D., Gentile, M. S., & Visalberghi, E. (2006). Are capuchin monkeys (Cebus apella) inequity averse? *Proceedings of the Royal Society B: Biological Sciences*, *273*(1591), 1223–1228. http://doi.org/10.1098/rspb.2005.3433

Engelmann, J. M., Clift, J. B., Herrmann, E., & Tomasello, M. (2017). Social disappointment explains chimpanzees' behaviour in the inequity aversion task. *Proceedings of the Royal*

Society B: Biological Sciences, *284*(1861). http://doi.org/10.1098/rspb.2017.1502

Engelmann, J. M., & Tomasello, M. (2017). Prosociality and Morality in Children and Chimpanzees. In C. Helwig (Ed.), *New Perspectives on Moral Development* (pp. 55–72). New York: Routledge. http://doi.org/10.4324/9781315642758

Engelmann, J. M., & Tomasello, M. (2019). Children's Sense of Fairness as Equal Respect. *Trends in Cognitive Sciences*, 1–10. http://doi.org/10.1016/j.tics.2019.03.001

Enoch, D. (2018). Why I Am an Objectivist about Ethisc (And Why You Are, Too). In R. Shafer-Landau (Ed.), *The Ethical Life: Fundamental Readings in Ethics and Moral Problems* (Fourth, pp. 208–221). New York: Oxford University Press.

Falkenstein, L. (2010). Étienne Bonnot de Condillac. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2010).

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. http://doi.org/10.1016/S1090-5138(04)00005-4

Fichte, J. G. (2000). *Foundations of Natural Right*. (F. Neuhouser, Ed.)*Cambridge texts in the history of philosophy*. Cambridge: Cambridge University Press.

FitzPatrick, W. (2016). Morality and Evolutionary Biology. In *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)* (pp. 1–30).

Flack, J. C., & de Waal, F. B. M. (2000). a Building Block of Morality. *Journal of Consciousness Studies*, *7*(1–2), 67–77.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, *6*, 347–369.

Friedman, M. (1989). Friendship and moral growth. *The Journal of Value Inquiry*, *23*(1), 3–13. http://doi.org/10.1007/BF00138682

Gallagher, S. (2005). *How the body shapes the mind*. Oxford University Press.

Gauthier, D. (1986). *Morals By Agreement*. Oxford: Oxford University Press.

Gibbard, A. (1982). Human Evolution and the Sense of Justice. *Midwest Studies in Philosophy*, *7*(1), 31–46. http://doi.org/10.1111/j.1475-4975.1982.tb00082.x

Gibbard, A. (1989). Communities of judgment. *Social Philosophy and Policy*, *7*(1), 175–189. http://doi.org/10.1017/S0265052500001072

Ginsborg, H. (2014). Kant's Aesthetics and Teleology. *The Stanford Encyclopedia of Philosophy*, *Fall 2014*.

Goldman, A. I. (1992a). Empathy, mind, and morals. *Proceedings and Addresses of the American Philosophical Association*, *66*(3), 17–41.

Goldman, A. I. (1992b). In Defense of the Simulation Theory. *Mind & Language*, *7*(1–2), 104–119. http://doi.org/10.1111/j.1468-0017.1992.tb00200.x

Gómez, J. C. (1998). Are apes persons? The Case for Primate Intersubjectivity. *Etica & Animali*, *9*, 51–63.

Gomila, A. (1994). Evolución y lenguaje. In *La Mente. Enciclopedia Iberoamericana de filosofía*. Trotta/CSIC.

Gomila, A. (2001a). La perspectiva de segunda persona: mecanismos mentales de la intersubjetividad. *Contrastes*, *6*, 65–86. http://doi.org/10.24310/Contrastescontrastes.v0i0.1448

Gomila, A. (2001b). Personas primates. In J. M. García Gómez-Heras (Ed.), *Ética del medio ambiente: problemas, perspectivas, historia* (pp. 191–206). Madrid: Tecnos.

Gomila, A. (2002). La perspectiva de segunda persona de la atribución mental. *Azafea: Revista de Filosofía*, *1*, 123–138.

Gomila, A. (2008). La relevancia moral de la perspectiva de segunda persona. In *D. Pérez y L. Fenández, L. (Eds.), Cuestiones filosóficas: ensayos en honor de Eduardo Rabossi* (pp. 493–

510).

Gomila, A. (2012). A Naturalistic Defense of "Human Only" Moral Subjects. *Dilemata*, (9), 69–73.

Gomila, A. (2015). Emociones en segunda persona. *X Boletín de Estudios de Filosofía Y Cultura Manuel Mindán*, *10*, 37–50.

Gomila, A., & Amengual, A. (2009). Moral Emotions for Autonomous Agents. In J. Vallverdú (Ed.), *Handbook of research on synthetic emotions and sociable robotics: New applications in affective computing and artificial intelligence* (pp. 166–180). IGI Global.

Gomila, A., & Pérez, D. (2017). Lo que la segunda persona no es. In D. Pérez & D. Lawer (Eds.), *La segunda persona y las emociones* (Vol. 0, pp. 275–297). Buenos Aires: Editorial SADAF.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*(3), 1339–1366. http://doi.org/10.1016/j.cognition.2007.06.007

Goodwin, G. P., & Darley, J. M. (2010). The Perceived Objectivity of Ethical Beliefs: Psychological Findings and Implications for Public Policy. *Review of Philosophy and Psychology*, *1*(2), 161–188. http://doi.org/10.1007/s13164-009-0013-4

Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, *48*(1), 250–256. http://doi.org/10.1016/j.jesp.2011.08.006

Gordon, R. M. (1992). The Simulation Theory: Objections and Misconceptions. *Mind & Language*, *7*(1–2), 11–34. http://doi.org/10.1111/j.1468-0017.1992.tb00195.x

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, *205*(1161), 581–598. http://doi.org/10.1098/rspb.1979.0086

Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York:

The Penguin Press. http://doi.org/10.15713/ins.mmj.3

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, *3*(1), 65–72. http://doi.org/10.1111/j.1745-6916.2008.00063.x

Hamilton, W. D. (1964). The Genetical Evolution of Social Behavior. *Journal of Theoretical Biology*, *7*(1), 17–52. http://doi.org/10.1016/0022-5193(64)90039-6

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). *Emotional Contagion*. *Current Directions in Psychological Science (Wiley-Blackwell)* (Vol. 2). Cambridge: Cambridge University Press. http://doi.org/10.1111/1467-8721.ep10770953

Hatfield, E., Rapson, R. L., & Le, Y.-C. L. (2009). Emotional contagion and empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 19–31). MIT Press.

Hegel, G. W. F. (1979). *Phenomenology of the Spirit*. (A. V. Miller, Ed.). Oxford: Oxford University Press.

Held, V. (2006). *The Ethics of Care: Personal, political, and global*. Oxford: Oxford University Press.

Helm, B. (2013). Friendship. In *The Stanford Encyclopedia of Philosophy* (pp. 1–38).

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815. http://doi.org/10.1017/S0140525X05000142

Heyes, C. (2018). Empathy is not in our genes. *Neuroscience and Biobehavioral Reviews*, *95*(October), 499–507. http://doi.org/10.1016/j.neubiorev.2018.11.001

Hobbes, T. (1651). *Leviathan*. (N. Malcom, Ed.). Oxford: Oxford University Press.

Hoffman, M. L. (2001). *Empathy and Moral Development: Implications for Caring and Justice*. *Review Zora Raboteg-Šarić Contemporary Sociology* (Vol. 30).

http://doi.org/10.2307/3089337

Hume, D. (1740). *A Treatise of Human Nature*. (L. A. Selby-Bigge, Ed.). Oxford: Oxford University Press.

Iacoboni, M. (2009). Imitation, Empathy, and Mirror Neurons. *Annual Review of Psychology*, *60*(1), 653–670. http://doi.org/10.1146/annurev.psych.60.110707.163604

Isen, A. M., & Levin, P. F. (2017). Effect of feeling good on helping: Cookies and kindness. *Social Psychology in Natural Settings: A Reader in Field Experimentation*, *21*(3), 97–106. http://doi.org/10.4324/9781315129747

Isern-Mas, C., & Gomila, A. (2018). Externalization is common to all value judgments, and norms are motivating because of their intersubjective grounding. *Behavioral and Brain Sciences*, *41*. http://doi.org/10.1017/S0140525X18000092

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage*, *24*(3), 771–779. http://doi.org/10.1016/j.neuroimage.2004.09.006

James, R., & Blair, R. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders*, *26*(5), 571–579. http://doi.org/10.1007/BF02172277

Jensen, K. (2016). Prosociality. *Current Biology*, *26*(16), R748–R752. http://doi.org/10.1016/j.cub.2016.07.025

Jeske, D. (2014). Special obligations. In *Stanford Encyclopedia of Philosophy* (pp. 1–22). http://doi.org/10.1111/1467-9973.00225

Joyce, R. (2006). *The Evolution of Morality*. Cambridge: MIT Press.

Kant, I. (1784). Moral Philosophy: Collins's lecture notes. In P. Heath & J. B. Schneewind (Eds.), *Lectures on Ethics: The Cambridge edition of the works of Immanuel Kant* (pp. 37–222). Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781107049512.004

Kant, I. (1785). Groundwork of the Metaphysics of Morals. In M. J. Gregor & A. W. Wood (Eds.), *Practical Philosophy: The Cambridge edition of the works of Immanuel Kant* (pp. 37–108). Cambridge: Cambridge University Press.

Kant, I. (1788). Critique of Practical Reason. In M. J. Gregor & A. W. Wood (Eds.), *Practical Philosophy: The Cambridge edition of the works of Immanuel Kant* (pp. 133–272). Cambridge: Cambridge University Press.

Kant, I. (1797). The Metaphysics of Morals. In M. J. Gregor & A. W. Wood (Eds.), *Practical Philosophy: The Cambridge edition of the works of Immanuel Kant* (pp. 353–604). Cambridge: Cambridge University Press.

Kant, I. (1987). *Critique of judgment*. (W. Pluhar, Ed.). Indianapolis: Hackett.

Korsgaard, C. M. (2010). Reflections on the evolution of morality. *The Amherst Lecture in Philosophy*, *5*, 1–29.

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, *37*(9), 1539–1553. http://doi.org/10.1037/0022-3514.37.9.1539

Levi, P. (2017). *The Drowned and the Saved*. (R. Rosenthal, Ed.). New York: Simon & Schuster Paperback.

Lévinas, E. (1969). *Totality and Infinity: An Essay or Exteriority*. (A. Lingis, Ed.). Pittsburgh: Duquesne University Press.

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, *21*(2), 153–174.

Lim, A., & Okuno, H. G. (2015). A Recipe for Empathy: Integrating the Mirror System, Insula, Somatosensory Cortex and Motherese. *International Journal of Social Robotics*, *7*(1), 35–49. http://doi.org/10.1007/s12369-014-0262-y

Maibom, H. L. (2009). Feeling for Others: Empathy, Sympathy, and Morality. *Inquiry*, *52*(December 2014), 483–499. http://doi.org/10.1080/00201740903302626

Martínez, M. (2003). La evolución del altruismo. *Revista Colombiana de Filosofía de La Ciencia*, *4*(9), 27–42.

Masto, M. (2015). Empathy and its role in morality. *Southern Journal of Philosophy*, *53*(1), 74–96. http://doi.org/10.1111/sjp.12097

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 491–500. http://doi.org/10.1098/rstb.2002.1261

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*(4), 193–197. http://doi.org/10.1097/01.wnn.0000191292.17964.bb

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*(1), 33–43. http://doi.org/10.1037/0022-3514.81.1.33

Monsó, S. (2015). Empathy and morality in behaviour readers. *Biology and Philosophy*, *30*(5), 671–690. http://doi.org/10.1007/s10539-015-9495-x

Nagel, T. (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

Nagel, T. (1986). *The View From Nowhere*. New York: Oxford University Press.

Noddings, N. (2010). Moral education and caring. *Theory and Research in Education*, *8*(2), 145–151. http://doi.org/10.1177/1477878510368617

Okasha, S. (2013). Biological Altruism. In *Stanford Encyclopedia of Philosophy*.

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in Virtual Agents and

Robots. *ACM Transactions on Interactive Intelligent Systems*, *7*(3), 1–40. http://doi.org/10.1145/2912150

Parfit, D. (1984). *Rasons and Persons*. Oxford: Oxford University Press.

Pérez-Manrique, A., & Gomila, A. (2017). The comparative study of empathy: sympathetic concern and empathic perspective-taking in non-human animals. *Biological Reviews*. http://doi.org/10.1111/brv.12342

Pettigrew, T. F. (1998). Intergroup Contact Theory. *Annual Review of Psychology*, *49*(1), 65–85. http://doi.org/10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. http://doi.org/10.1037/0022-3514.90.5.751

Pierce, J., & Bekoff, M. (2012). Wild justice redux: What we know about social justice in animals and why it matters. *Social Justice Research*, *25*(2), 122–139. http://doi.org/10.1007/s11211-012-0154-y

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciencesav*, *25*, 1–72. http://doi.org/10.1017/S0140525X02000018

Prinz, J. (2011). Is Empathy Necessary for Morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 211–229). Oxford University Press.

Rachels, J., & Rachels, S. (2015). *The elements of moral philosophy* (8th ed.). New York: McGraw-Hill. http://doi.org/10.1017/CBO9781139519373.012

Rai, T. S., & Holyoak, K. J. (2013). Exposure to moral relativism compromises moral behavior. *Journal of Experimental Social Psychology*, *49*(6), 995–1001. http://doi.org/10.1016/j.jesp.2013.06.008

Railton, P. (1984). Alienation, Consequentialism, and the Demands of Morality. *Philosophy*

& Public Affairs, *13*(2), 134–171.

Railton, P. (1986). Moral Realism. *The Philsophical Review*, *95*(2), 163–207. http://doi.org/10.1093/oxfordhb/9780195325911.003.0002

Railton, P. (2016). Moral Learning: Why learning? Why moral? And why now? *Cognition*, *167*, 172–190. http://doi.org/10.1016/j.cognition.2016.08.015

Rhodes, M., & Chalik, L. (2013). Social Categories as Markers of Intrinsic Interpersonal Obligations. *Psychological Science*, *24*(6), 999–1006. http://doi.org/10.1177/0956797612466267

Ricoeur, P. (1954). Sympathie et respect: phénoménologie et éthique de la seconde personne. *Revue de Métaphysique et de Morale*, *59*(4), 380–397.

Riis, J., Simmons, J. P., & Goodwin, G. P. (2008). Preferences for Enhancement Pharmaceuticals: The Reluctance to Enhance Fundamental Traits. *Journal of Consumer Research*, *35*(3), 495–508. http://doi.org/10.1086/588746

Roma, P. G., Silberberg, A., Ruggiero, A. M., & Suomi, S. J. (2006). Capuchin monkeys, inequity aversion, and the frustration effect. *Journal of Comparative Psychology*, *120*(1), 67–73. http://doi.org/10.1037/0735-7036.120.1.67

Rosas, A. (2005). La moral y sus sombras: La racionalidad instrumental y la evolución de las normas de equidad. *CRÍTICA. Revista Hispanoamericana de Filosofía*, *37*(110), 79–104.

Rosas, A. (2013). Rationality and Deceit: Why Rational Egoism Cannot Make Us Moral. In B. Musschenga & A. van Harskamp (Eds.), *What Makes Us Moral? On the capacities and conditions for being moral* (p. 360). New York: Springer. http://doi.org/10.1007/978-94-007-6343-2

Rosati, C. (2016). Moral motivation. In *The Stanford Encyclopedia of Philosophy*.

Roskies, A. L. (2003). Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy" [1]. *Philosophical Psychology*, *16*(1), 51–66. http://doi.org/10.1080/0951508032000067743

Roskies, A. L. (2011). A Puzzle about Empathy. *Emotion Review*, *3*(3), 278–280. http://doi.org/10.1177/1754073911402395

Roughley, N. (2018). From shared intentionality to moral obligation? Some worries. *Philosophical Psychology*, *31*(5), 736–754. http://doi.org/10.1080/09515089.2018.1486610

Rousseau, J.-J. (1762). *The Social Contract*. (C. Betts, Ed.). New York: Oxford University Press. http://doi.org/10.1023/B:ELEC.0000045977.47306.16

Rowlands, M. (2012a). ¿Pueden los animales ser morales ? *Dilemata*, *9*(1989–7022), 1–32.

Rowlands, M. (2012b). Can Animals be Moral ? *Dilemata*, *4*(9), 1–32.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, *20*(4), 523–528. http://doi.org/10.1111/j.1467-9280.2009.02326.x

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind and Language*, *26*(4), 482–505. http://doi.org/10.1111/j.1468-0017.2011.01428.x

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*, 393–462.

Seyfarth, R. M., & Cheney, D. L. (2012). The Evolutionary Origins of Friendship. *Annual Review of Psychology*, *63*, 153–177. http://doi.org/10.1146/annurev-psych-120710-100337

Sheskin, M., Ashayeri, K., Skerry, A., & Santos, L. R. (2014). Capuchin monkeys (Cebus apella) fail to show inequality aversion in a no-cost situation. *Evolution and Human Behavior*, *35*(2), 80–88. http://doi.org/10.1016/j.evolhumbehav.2013.10.004

Sie, M. (2014). Self-Knowledge and the Minimal Conditions of Responsibility: A Traffic-

Participation View on Human (Moral) Agency. *Journal of Value Inquiry*, *48*(2), 271–291. http://doi.org/10.1007/s10790-014-9424-2

Sie, M. (2015). Moral Hypocrisy and Acting for Reasons: How Moralizing Can Invite Self-Deception. *Ethical Theory and Moral Practice*, *18*(2), 223–235. http://doi.org/10.1007/s10677-015-9574-8

Silberberg, A., Crescimbene, L., Addessi, E., Anderson, J. R., & Visalberghi, E. (2009). Does inequity aversion depend on a frustration effect? A test with capuchin monkeys (Cebus apella). *Animal Cognition*, *12*(3), 505–509. http://doi.org/10.1007/s10071-009-0211-6

Slote, M. (1999). Caring versus the Philosophers. *Philosophy of Education Archive*, 25–35.

Smith, A. (1759). *The Theory of Moral Sentiments*. (S. M. Soares, Ed.). São Paulo: MetaLibri.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishing Ltd.

Sober, E., & Wilson, D. S. (1998). *Unto others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press.

Stanford, P. K. (2018). The Difference Between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation. *Behavioral and Brain Sciences*, *41*, 1–57. http://doi.org/10.1017/S0140525X17001911

Stocker, M. (1976). The Schizophrenia of Modern Ethical Theories. *Journal of Philosophy*, *73*(14), 453–466. http://doi.org/10.1074/jbc.M112.418400

Strawson, P. F. (1974). Freedom and resentment. In P. F. Strawson (Ed.), *Freedom and Resentment and other Essays* (pp. 1–28). Abingdon: Routledge.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171. http://doi.org/10.1016/j.cognition.2013.12.005

Strohminger, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological*

Science, *26*(9), 1469–1479. http://doi.org/10.1177/0956797615592381

Stueber, K. (2014). Empathy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 201).

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology*, *58*(1), 345–372. http://doi.org/10.1146/annurev.psych.56.091103.070145

Thirioux, B., Mercier, M. R., Blanke, O., & Berthoz, A. (2014). The cognitive and neural time course of empathy and sympathy: An electrical neuroimaging study on self-other interaction. *Neuroscience*, *267*, 286–306. http://doi.org/10.1016/j.neuroscience.2014.02.024

Tobia, K. P. (2015). Personal identity and the Phineas Gage effect. *Analysis*, *75*(3), 396–405. http://doi.org/10.1093/analys/anv041

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2016). *A natural history of human morality*. Cambridge: Harvard University Press.

Tomasello, M. (2018). Response to commentators. *Philosophical Psychology*, *31*(5), 817–829. http://doi.org/10.1080/09515089.2018.1486604

Tomasello, M., & Vaish, A. (2013). Origins of Human Cooperation and Morality. *Annual Review of Psychology*, *64*(1), 231–255.

Trevarthen, C. (1977). Descriptive analyses of infant communication behavior. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: The Loch Lomond Symposium* (pp. 227–70). Academic Press.

Trevarthen, C. (1980). The foundations of intersubjectivity: Development of interpersonal and cooperative understanding in infants. In D. Olson (Ed.), *The social foundations of* *language and thought: Essays in honor of J. S. Bruner* (pp. 316–42). Norton.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. http://doi.org/10.1086/406755

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13. http://doi.org/10.1016/S0002-7138(09)62273-1

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

van Baaren, R. B., Decety, J., Dijksterhuis, A., van der Leij, A., & van Leeuwen, M. L. (2009). Being imitated: consequences of nonconsciously showing empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 31–42). Cambridge: MIT Press.

Vincent, S., Ring, R., & Andrews, K. (2019). Normative Practices of Other Animals. In A. Zimmerman, K. Jones, & M. Timmons (Eds.), *The Routlege Handbook of Moral Epistemology* (pp. 57–84). New York: Routledge.

Wallace, R. J. (2007). Reasons, Relations, and Commands: Reflections on Darwall. *Ethics*, *118*(1), 24–36. http://doi.org/10.1086/522016

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303. http://doi.org/10.1126/science.1121448

Wilkinson, G. S. (1990). Food sharing in vampire bats. *Scientific American*, *262*(2), 76–83.

Williams, B. (1981). Persons, character and morality. In A. van der Leij (Ed.), *Moral Luck* (pp. 1–19). Cambridge: Cambridge University Press.

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, *50*(2), 314–321.

http://doi.org/10.1037/0022-3514.50.2.314

Wolf, S. (2012). “One Thought Too Many”: Love, Morality, and the Ordering of Commitment. In U. Heuer & G. Lang (Eds.), *Luck, Value, and Commitment. Themes from the Ethics of Bernard Williams* (pp. 71–92). Oxford: Oxford University Press.

Young, L., & Durwin, A. J. (2013). Moral realism as moral motivation: The impact of meta-ethics on everyday decision-making. *Journal of Experimental Social Psychology*, *49*(2), 302–306. http://doi.org/10.1016/j.jesp.2012.11.013

Zangwill, N. (2014). Aesthetic Judgment. *The Stanford Encyclopedia of Philosophy*, *Fall 2014*.