ACTA
PHILOSOPHICA
FENNICA

VOL. 94 2018

Action, Value and Metaphysics

Proceedings of the Philosophical Society of Finland Colloquium 2018

Edited by
JAAKKO KUORIKOSKI & TEEMU TOPPINEN

SOCIETAS PHILOSOPHICA FENNICA

ACTA PHILOSOPHICA FENNICA

Acta Philosophica Fennica is published by the Philosophical Society of Finland. Since the inception of the series in 1935 it has been the forum for much of the best philosophical work in Finland. In 1968–1981 it was distributed by the North-Holland Publishing Company (Amsterdam), and since 1981 by the Academic Bookstore and Bookstore Tiedekirja (Helsinki).

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Information for Authors

The Acta series publishes shorter and longer monographs as well as collections of articles in all parts of philosophy. Authors should send their contributions to

Acta Philosophica Fennica,
Department of Philosophy,
P.O. Box 24 (Unioninkatu 40 A),
FI-00014 University of Helsinki,
Finland.

All philosophical traditions and all types of philosophy fall within the intended scope of the Acta Philosophica Fennica. It follows from its traditional character, however, that special consideration is given to the work of Finnish philosophers and to papers and monographs inspired by their contributions to philosophy.

Subscription Information

Permanent subscriptions can be placed directly with Bookstore Tiedekirja, Snellmaninkatu 13, FI-00170 Helsinki, Finland, tel. +358–9–635 177, email: tiedekirja@tsv.fi, www.tiedekirja.fi. Other orders can be placed online with Bookstore Tiedekirja, www.tiedekirja.fi.

ACTA PHILOSOPHICA FENNICA

Vol. 94

Action, Value and Metaphysics

Proceedings of the Philosophical Society of Finland Colloquium 2018

Edited by
JAAKKO KUORIKOSKI & TEEMU TOPPINEN

HELSINKI 2018

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

ISBN 978−951−9264−89−9
ISSN 0355-1792

Juvenes Print
Helsinki 2018

Preface

The Philosophical Society of Finland is one of the country's oldest learned societies with practices steeped in traditions, formulated over the years by a number of giants in the field. Yet sometimes change is for the better and so the board, recently undergone a generational shift, decided to remodel the yearly society colloquium. The objective was an open, general, and rigorously refereed conference in which philosophers could disseminate and discuss their best work. Although the mission of the society is to foster Finnish philosophy – in universities and in society at large - the meaning of what is to be Finnish has been and is changing. Although the philosophical community in Finland has always been international in that Finnish philosophers have always been visible participants in the international philosophical community, the philosophical community working within Finland has recently become much more international – and hence much less Finnish or Swedish speaking. Therefore it was decided that the conference presentations could also be given in English, not only to attract contributors beyond our borders, but most of all to better reflect and serve the diversifying Finnish philosophy. The new colloquium was held in Helsinki on 11-12 January 2018.

This collection is a result of an open call for papers based on talks given at the colloquium. Although we as editors did not aim at any thematic statements or even coherence, and simply let quality decide over subject matter, the collection ended up nicely representing many of the strongest subfields in the present philosophical landscape in Finland.

Säde Hormio explores different ways in which organizational design can lead to individual ignorance, sometimes creating conditions for culpable ignorance. Her work is therefore an excellent and timely example of research on *collective responsibility*, a line of research which has flourished as part of a broader interest in social ontology. The same broad interest in understanding the foundations of social action has also meant that *action theory* in general has been very much in focus in Finland. Harry Alanen offers us an assessment of Davidson's understanding of Aristoteles' account of action and thus helps us better locate current philosophical projects related to action within a broader historical outlook.

Metaethics has in recent years been one of the most vibrant areas of research in Finnish philosophy, broadly construed, and Antti Sneitz's contribution gives us yet another original take on the nature of human good. Sneitz redevelops Boyd's brand of moral realism, based on the idea of homeostatic property clusters, in a new direction by linking the view to Aristotle's account of virtues.

Frank Martela explores the boundary between empirical psychology of well-being and philosophical accounts of objective value. He argues that certain values, such as happiness, morality, contribution and authenticity, can be arguably considered as self-justifying on empirical grounds. Martela's work is part of a resurgence of thinking about *well-being, happiness and meaning* in philosophy and also exemplifies the porousness of the boundaries between philosophy and empirical sciences.

Finally, Jani Hakkarainen and Markku Keinänen continue to build on the Finnish tradition of metaphysics in general, and *tropical realism* in particular. Hakkarainen provides a formal ontological account of tropes as they are found in the strong nuclear theory of tropes and substances, and argues that the full fundamental ontological form of every trope is to be a strongly rigidly or generically dependent individual entity that is a simple part. Keinänen also builds on the strong nuclear theory by developing a thus far missing reductive but non-eliminativist account of relational entities.

We would finally like to thank the referees, without whose efforts this collection would not exist.

The Editors

Table of Contents

Culpable Ignorance in a Collective Setting

SÄDE HORMIO

1. Introduction

Ignorance is traditionally seen as an excuse for blame. While there has been a growing interest in ignorance and individual responsibility,[1] literature on collective ignorance and what it means for responsibility has so far been quite thin on the ground. This paper hopes to bridge some of the gaps in the literature by exploring ways in which organisational practices can affect the knowledge we have about the causes and effects of our actions.

My concern is with the *epistemic condition of responsibility*, of acting under ignorance, not with *epistemic responsibility* (i.e. questions regarding what you ought or ought not to believe). If one does not know what one is involved in, and cannot be reasonably expected to know either, then one cannot be praiseworthy or blameworthy. But what about the collective? Organisational practices can affect the knowledge we have about the causes and effects of our actions. Can an organization be blameworthy when an individual acts under ignorance?

There are three interlinked issues that I will explore in this paper. The first is about discovering what different types of ignorance can be found in organisations. The second describes how sometimes organisational design creates ignorance even without anyone trying deliberately to mislead anyone. The third concerns how organisations can be responsible for an individual's culpable ignorance. I explore these questions mainly through fragmentation of information and suppression of knowledge.

[1] See for example the collection of papers in Peels 2016 and Robichaud and Wieland 2017.

I will discuss responsibility at two levels: at the level of the organisation and at the level of the individual members of the organisation. Organisational ignorance can be ignorance about facts or it can arise from the suppression of knowledge. My discussion of members in an organisation is focused on the "regular employees" in a large organization who are subjected to the standard effects of compartmentalization of information that is necessary in large organised collectives. The assessment of excusing or culpability of individual ignorance is therefore in most cases aimed at this "general" level, not at individuals that occupy special positions in the organisation with privileged access to information, such as directors. The assessment of their individual responsibility might vary a lot from the responsibility of the average employee, depending on circumstances.

This paper proceeds as follows. I begin by discussing ignorance from the point of view of the collective. Section 2 introduces types of ignorance about knowledge and facts, while section 3 discusses ignorance arising from suppression of knowledge. Section 4 looks at organisational design and how fragmentation of information can lead to ignorance. I then turn my attention to individual members of collectives in section 5, where I discuss culpable ignorance in collective setting and suggest that in some cases, the blame for culpable ignorance should be directed at the level of the organisation.

2. Ignorance about knowledge and facts

The role of ignorance in society is not just passive and negative. Often ignorance is simply unavoidable or neutral. Noone can know everything, and not all information is relevant for all. Ignorance is also an indispensable element in many social relations and structures (Moore and Tumin 1949). Ignorance is not a static state of affairs. In the course of time, unknowns are transformed into new knowledge, while at the organisational level some old knowledge is forgotten and replaced with ignorance (Roberts 2013, 218). Science challenges the existing body of knowledge and is at the centre of most of our efforts of turning unknowns and ignorance into knowledge.

This section introduces issues related to ignorance about knowledge and facts, while in the next section ignorance arising from the suppression of knowledge is discussed. I will use Joanne Roberts's (2013) work on organisational ignorance as my starting point. [2] While the concern in Roberts's original work is not with normative questions or moral responsibility, I will apply her categories in this way.[3]

Ignorance about knowledge and facts can be divided into three subcategories. *Knowable recognised unknowns* are something that could be found out given the right resources and

[2] Roberts (2013) divides ignorance into three key sources: ignorance arising from the absence of knowledge, ignorance about knowledge, and ignorance arising from the suppression of knowledge. This paper concentrates on the latter two, but I will briefly describe what is involved in the first category before setting it aside. *Known unknowns* are things outside the limits of our knowledge, things we know that we do not know. *Unknown unknowns* are beyond anticipation, a total lack of knowledge: something we are not even aware of being ignorant about at a specific point in time, so they cannot be directly investigated. Examples of both can be found in astrophysics and the current knowledge of our solar system. Known unknowns refers to a state of ignorance at a specific point in time in an organizational context, it is an awareness that certain knowledge is not in fact known by the organisation or its members. Organizational known unknowns can drive research and development, leading to innovations. They are also something that needs to be taken into account when the organisation is involved with research and development that carries high risks. (Roberts 2013, 217-221). Due to this, there could be some cases of collective responsibility linked to known unknowns, but even in them the potential blame would not be likely linked to culpable ignorance. When it comes to unknown unknowns, undoubtedly advances in science will bring new knowledge that is currently outside the scope of our knowledge (and create new unknowns in the process), or even beyond our anticipation. However, as unknown unknowns are so completely outside our control, I do not find the type to be relevant for debates on blameworthy ignorance.

[3] I make no claim to offer an exhaustive list of all possible types of ignorance in an organisational setting, but I believe that Roberts's work offers a comprehensive typology. Management science is an interdisciplinary field where the goal is to find solutions to organizational problems and challenges, and her work draws from research into ignorance from a variety of fields, including sociology, politics, economics and philosophy. I therefore believe it to be a fertile ground to start mapping out normative framework on ignorance.

motivation. In Roberts's paper these recognised unknowns that are knowable (i.e. the ignorance about the facts could be overcome) are called "knowable known unknowns", but this term can be philosophically confusing due to the tautology involved. Therefore I am referring to this category as knowable recognised unknowns throughout. This ignorance could be overcome if the organisation wanted to, i.e. there is no missing science or technology standing in the way of obtaining knowledge about these facts.[4] Knowable recognised unknowns are either outside the focus of the organisation, resulting in lack of motivation to overcome the ignorance, or the costs and benefits do not add up to put the expenditure necessary into resources to overcome it. To give an example, an educational organisation working in India should be sensitive to issues to do with discrimination against Dalit students, while a similar organisation working in Argentina does not necessarily need to know anything about the issue.

Unknown knowns are things we do not know that we know, including tacit knowledge, meaning that this type of ignorance does not necessarily prevent the use of the knowledge. In fact it may underpin creativity in the form of intuition. In organizations unrecognized knowledge is often embedded in routines and collective practices, existence of which is usually exposed only in retrospect once it is lost, for example when people retire. Finally *errors* arise from inaccuracy, confusion, uncertainty or incompleteness. They are the things we think we know, but don't. The more complex an organization is, the more it is prone to errors. Individuals might either wrongly assess their level of competence, resulting in an error that has implications for the wider organisation, or there is a system failure due to a design fault in the organization. Changing environments increase the risk of organizational errors occurring. (Roberts 2013, 218-223).

I will soon discuss how knowable recognised unknowns and errors can raise issues of responsibility and blameworthiness in a collective setting. However, I will first argue that it is harder to see how unknown knowns could raise these issues. Tacit knowledge is difficult to articulate and sharing it is not a simple matter, in an organizational setting. While one

[4] In contrast to known unknowns mentioned in footnote 2.

could argue that it should be part of good managerial practise to try to ensure that as much as possible of the silent knowledge is passed on to the new employees as old ones retire, for example, the fact that it is hard to measure when this passing on has been successful (or not) belies that tacit knowledge is not a very fruitful basis for blame arising from ignorance. Of course, if an organisation makes these kind of practices difficult or impossible (for example, letting go the old staff before new staff starts, not asking for any handover notes etc.), it could be argued to be not just engaged in deficient investigation, but also involved in preventing subsequent discovery (categories that I will come to in section 5). Still, I find that genuine cases of culpable ignorance remain limited for this type of ignorance.

To put knowable recognised unknowns and errors in more concrete terms, I will introduce the tale of the exploding toasters. The example will help to illustrate responsibility issues related to ignorance in collective contexts, in this narrative in supply chain management. Supply chain management is a term with many usages, as it covers the product cycle from design of new products and services to the delivery of the finished product to the end customers (Lu and Swaminathan 2015). I will use the term here to refer to procurement of goods and to how companies manage outsourcing of the manufacturing of their products.

Violet works as an in-buyer for a large retail chain Sell-A-Lot with shops all over the world. She finalises a large toaster purchase order. Unbeknownst to Violet, the toaster company, Exciting Electronics, has very recently changed their manufacturer to cut costs. The decision to make changes to the supply chain was made at the last minute with regards to the upcoming peak sales period. As a consequence, the new toasters are manufactured in a rush by the new supplier and some have loose wires. Exciting Electronics sales representative Sharon is not aware of the problem and sells faulty toasters to Sell-A-Lot in good faith. Before long shop managers at Sell-A-Lot are flooded with angry phone calls from customers about their brand-new toasters sending off sparks. One unlucky soul has her house set on fire but lives to see another day. There need not exist any malicious intent; rather, were are looking at negligence. Nobody in the new factory supply-ing Exciting Electronics makes the wires loose on purpose and with evil intent, all faults are due to the too hectic manufacturing process.

Errors can raise questions of blameworthiness in many ways, the most serious ones being systematic errors. A systematic error of some sort must have taken place at Exciting Electronics for Sharon to be able to sell faulty toasters to a customer in the first place. It looks like a clear case of design fault in the organization if sales are allowed to go through on faulty products if the company knows about the fault. Systemic errors are blameworthy, as I will argue later on. In case of genuine errors, ones that are not due to negligent practises, there will be no blame (although there might be certain sense of agent-regret as I argue in section 5).

Knowable recognised unknowns were something that could be found out given the right resources and motivation. With this category, I think that we should keep in mind that organisations have less excuses of being ignorant of facts than individuals do. They can set aside money to fund research into a suitable course of action for them and have a group of experts dedicate their working hours to thinking through an issue from the organisation's point of view. If the required expertise cannot be found among their existing members, they can hire new staff or employ consultants. They can and should do this when new issues arise that affect their operating environment and future operations. Saying that, if something is defensibly outside the focus of the organisation, leading to lack of motivation to overcome the ignorance, I find that the resulting ignorance is not susceptible to blame. In the case of Exciting Electronics, this could be the environmental impact of hairdryers, for example, if they did not manufacture any such products.

Trickier variety of knowable recognised unknowns are cases where the costs and benefits do not justify the expenditure to attempt to overcome ignorance, so the decision to remain ignorant is made on a purely financial basis. I say trickier because cost and benefit analyses of any sort are riddled with normative assumptions. One board of directors in one company might deem something to be too costly, while another one would rule it to be a justifiable expenditure. If Sell-A-Lot or Exciting Electronics has decided that it is too costly to put in the resources necessary to adequately monitor their supply chain in relation to, say, the working conditions in the factories, then these working conditions remain knowable recognised unknowns to the organisations. Arguably this situation could (and should) be different, but such arguments need to be rooted in normative considerations outside simple cost and benefit analyses.

More broadly put, when it comes to knowable recognised unknowns, the issue of whether ignorance is blameworthy or excusing can only be settled in the context of deciding the adequate focus for the organisation. Following Tuomela (2007, 15) and Laitinen (2014, 218), this focus could be labelled as the *ethos* of the organisation, consisting of the central questions and practical matters that are vital to the purpose of the group (the group's *realm of concern*) and the answers it has collectively accepted to be its view (*intentional horizon*). Ethos thus covers the central goals and commitments of the organisation. The exploding toasters would be a case of knowable recognised unknowns if Sell-A-Lot has poor supply chain management and little motivation to invest properly in even basic product safety testing, let alone other corporate social responsibility measures. The harm caused by the faulty toasters could then be traced back to corporate policies and priorities in addition to the obvious fault on the manufacturing side. In other words, the ethos of the organisation could be argued to include negligence towards safety.

What is important for the topic at hand is that while the ethos determines a group's identity (and is part of what marks its continuation together with its historical and modal properties), it is in a state of flux (to what degree varies naturally a lot between each case). The ethos of a group is therefore not set in stone, as elements of it may change (and almost always do to some degree at least, especially when it comes to the large organisational-size groups). Examples are everywhere: corporations venture into new areas of production, political parties amalgamate new goals, the jurisdiction of local authorities change, university begins to offer courses in a new subject matter; it is easy to keep coming up with examples. Therefore, to simply state that some knowledge is currently outside the realm of concern of the ethos of an organisation does not, by itself, settle much in terms of responsibility and possible blameworthiness.

Because organisations are responsive to their environments and must regularly review their realm of concerns as well their intentional horizons, I argue that knowable recognised unknowns are always a normative matter. They could provide an interesting angle for political philosophy to look at questions related to what should fall within the ethos of the organisation. For example, how much and to what degree should a government be aware of the impacts of the actions of the banks it has chartered? Or looking at corporate responsibility, how far into their supply chain should a corporation look? However, ignorance about knowledge and facts is not the only category of organisational ignorance that is ripe for philosophical analysis on responsibility. I turn to ignorance that results from suppression of knowledge in the next section.

Before moving on, I will briefly address the question of what organisations can be said to know. I have so far discussed knowledge that the organisation does not have, but what about the knowledge an organisation can be argued to have? Imagine a spy ring of some kind, where the spies do not know the identity of the other spies or have access to the information the others have.[5] In my example, each spy has been assigned a code name and a secret phone with which to get in touch with the others. They have an assignment to complete where Aja knows the target, Katya the method, and Shea the time and the place. The person who set them up on the mission was involved in a car crash and lies in a coma in hospital somewhere. Shea has been instructed to invite the other two to come to the designated place at the designated time, Katya to bring the means, and Aja to put it in use towards the target. Between them, they have all the information necessary to successfully complete the assignment of the spy ring. However, can the spy ring as a collective be said to have knowledge about what the assignment is (before it is carried out)?

I suggest that the spy ring does have working knowledge about the assignment, as it is able to carry it out. If the spy

[5] I thank Jaakko Kuorikoski for suggesting using a spy ring as an example.

ring operates as part of some organised espionage group, it is the institution rather than the spy ring per se that knows what the assignment is.[6] However, if the person who is now in coma is some rogue agent (or some eccentric millionaire who engages in espionage as a hobby etc.), then the spy ring per se has the knowledge. Saying that, the knowledge of the spy ring is highly fragmented. Therefore the group knowledge is not robust at all: if one link was missing, they would not be able to achieve their goal. In this case, the spy ring knows *how* to perform the assignment, but not *that* the assignment is to *X*.[7] After they have carried it out, the spy ring understandably also knows more about the assignment - who was the target, where, and the means - as the knowledge of individual group members has come together in action. Knowing both how to do something and the details of what you are doing is clearly more robust group knowledge than knowing just the former.

By bringing in robustness, I am suggesting that group knowledge comes in degrees. Furthermore, I suggest that making the group knowledge more robust could in some cases be thought of as something that the collective should do. With highly fragmented information, the collective is taking a risk that the goal is not reached. In case of spy rings, this risk seems acceptable, as there are benefits to the arrangement: if Katya is captured, the secret mission is not revealed. However, in many everyday cases the information a collective has could be too fragmented and the associated risks not acceptable, like with the loose wires and internal practices of Exciting Electronics. In these cases, the ethos of the collective could be argued to be too negligent towards supply chain management. I will return to these issues in section 4, where I discuss organisational fragmentation of information.

3. Suppression of knowledge

This section introduces types of organisational ignorance that can arise from the suppression of knowledge. Withholding some important information, or the tendency to only com-

[6] I thank Deborah Tollefsen for pushing me on this point.

[7] I am indebted to an anonymous reviewer for making me draw up further distinctions on this point.

municate the positive news, is common among corporations and other large modern organisations. This section looks at different reasons for suppressing knowledge.

Taboos are socially constructed bans on certain types of knowledge deemed to be polluting. They can also be actively cultivated within organisations to influence the way its members behave, like a taboo about discussing bullying in the workplace. When knowledge is too painful to acknowledge, or it does not fit with one's worldview, it can be repressed or ignored, resulting in *denials*. Organisational denials can lead to ignoring evidence that contradicts the group decision for the sake of unanimity. This can be especially dangerous when encouraged by those in charge, as toleration for recklessness and dishonesty in practices has a tendency to spread. Denials can also be used strategically, like when a company encourages ignorance in their customers through misinformation campaigns. We talk of *secrecy* when knowledge is consciously suppressed by individuals or collectives. Pockets of ignorance can be deliberately created for power purposes. Some secrecy is essential (keeping trade secrets, for example) but there has to be a balance and an understanding of how much secrecy the stakeholders are willing to tolerate. *Privacy* is socially sanctioned secrecy and the right to privacy is enshrined in many laws and declarations. To build trust between an organisation and its members and stakeholders, it is important to recognise and protect privacy, for example, the customer data registry of a company. (Roberts 2013, 218-226).

I find that privacy could also be thought as a sub-category of secrecy. Privacy here refers to the privacy of the organisations' employees or customers. It is often about trust, the disability of an employee need not become common knowledge within the workplace, and not keeping customer data safe can by itself be a morally blameworthy act that can lead to the loss of those customers, or even to harm for those customers, depending on your line of business. Organisational ignorance always has power dimensions and the potential to be political in nature. To give an example, privacy has been in the news a lot lately with investigations into how Facebook has handled its users' data. Still, I will set this category aside for the rest of the paper, as my concern is with cases where organisations can make or keep individuals ignorant about issues they arguably should not be ignorant about, not with issues to do with mismanaging data.

It seems clear that many instances of ignorance arising from the other forms of suppression of knowledge can be amenable to blameworthiness. Think of some corporation that is involved in practices that are questionable, but is not communicating this to its employees or other stakeholders. A recent real-life example of this is how many employees of Google were taken aback when they found out that the company is involved in developing algorithms for military use. Google's participation in Project Maven with the Pentagon prompted thousands of its employees to sign an open letter urging Google to not be involved in developing military Artificial Intelligence. When it comes to suppression of knowledge, an organisation has introduced a condition (through denial, secrecy, or taboo) – or failed to remove it – which made it difficult for employees to acquire true belief about the wrongness of being involved in some particular collective action. Depending on the actual circumstances, the employees' ignorance could be culpable if it is due to deficient inference, or excusable if the organisational barrier for acquiring the knowledge is too high, but I will return to this section 5.

Secrecy is very common in organisations. Although we live in the information age, secrecy is still a very prevalent component of our societies. For example, Galison (2008, 38) describes how "we are living in a climate of augmented secrecy" today, with the number of classified document pages outnumbering the amount of open literature entering the public libraries and archives each year in the U.S.[8] Secrecy naturally has a large role to play in organisations such as intelligence services, but it is an ubiquitous part of the corporate world also. Secrecy has positive features for an organisation, like affording more freedom in negotiating difficult situations in politics, or giving a competitive advantage to a corporation by ensuring first-mover advantage in a new product area (Dufresne and Offstein 2008). Still, secrecy can also be used as an excuse. Coming back to the example I have

[8] Galison (2008, 37-39) attributes this rise of modern censorship mainly to the infrastructure created after the Second World War around nuclear science and intelligence services.

been using, if Exciting Electronics had a policy in place where all information about suppliers were classified as trade secrets, it would be impossible for Sell-A-Lot to know where the products it sells originate from. If the ethos of Sell-A-Lot includes ensuring safe working conditions in its supply chain, this kind of non-transparency from the part of Exciting Electronics would be unacceptable for them, resulting in them procuring their toasters elsewhere.

While genuine trade secrets are one thing, like Google's algorithm or the closely guarded recipe of Coca-Cola, corporations can use secrecy simply to avoid awkward questions about their products or their supply chain without acceptable reasons for their secrecy. The acceptability has to be linked to the kind of reasons given for the secrecy: would they stand the scrutiny of objective outsiders? If there are no acceptable reasons for secrecy, I find that non-transparent business practices are not justifiable, at least in the modern world, where concerns about the treatment of workers in global supply chains or the environmental impacts of rampant consumerism have become widely known.

The problem with organisational secrecy is of course that often we do not know what we should be concerned about, at least until a whistle-blower alerts us to the facts. Still, there are areas that are widely known to be riddled with problems, like the conditions under which many of our clothes or consumer electronics are produced under. Therefore, unless the corporation is forthcoming about the way it handles the problematic areas in its supply chain, there is a high likelihood that they have a thing or two to hide. To use Exciting Electronics again, it could be the case that the top management were aware of the faulty toasters, but decided to deny this to protect profits in the short term. Maybe the quantities already delivered to foreign retailers were large, and costs of recalling products were deemed too great, because faults were found in only a few toasters so far, so they decided to gamble. Let's say Bianca is a conscientious middle-level manager supervising Sharon's department. She was aware that their toaster manufacturer was changed and wanted to make sure that the new products were tested for safety. Bianca enquired after the results from the top management, who denied knowledge of any faults. Bianca gave Sharon's team the go ahead to sell the toasters to Sell-A-Lot and other domestic retailers. Any blame should be directed at the top management in this case, or the organisation itself, depending on the details.

Denials, on the other hand, can take form of flat-out lies or more subtle agnotology, where the goal is to create misinformation (Proctor and Schiebinger 2008). I have argued elsewhere that an organisation can be blameworthy for distorting public debate through strategic denials (Hormio 2017), but here I want to address only ignorance within organisations, so I will give an example of internal denials. The loose wires could bring back painful issues for Exciting Electronics, so the information could be repressed or ignored by those who are privy to the manufacturing problems. Say that Exciting Electronics used to own its own factories, but changed its business models some years back in order to become more competitive with its prices. As part of this process, they closed down their own factories in Europe and outsourced their manufacturing to countries with cheap labour and laxer regulations. This decision was far from unanimous at the board level and raised a lot of debate at other levels of the organisation too, let alone among those who lost their jobs at the closed-down factories. The loose wires are therefore a painful reminder of the costs involved to those who really pushed for the outsourcing, thus motivating denials. These kind of denials are blameworthy, despite the psychological backstory. Saying that, there are also circumstances where organisational denials are clearly defendable, in the sense that objective outsiders would be likely to agree with the need for such measures. For example, national security concerns could provide such acceptable reasons for an organisational denial.

Coming to the level of individual members of an organisation for a moment, we might also not want to know. Maybe Bianca had an inclination as an experienced supply chain manager that something could be wrong, but did not seek to find answers. We might even actively avoid finding out about the consequences of our actions and choices. While I was working for an NGO that campaigned on ethical issues in global supply chains, an acquaintance once told me to never tell her anything bad about the multinational corporation she was working for, as she wanted to continue working for them. Although it was meant partly as a joke, this kind of attitude is typical of wilful ignorance, where there "is a self-interested reason for evading moral knowledge that might require one to rethink one's way of life" (Isaacs 2011, 162). Corporations with market shares to protect can actively play into this, and thus we get meat packages with pictures of happy farm animals grazing on a green pasture, instead of pictures about the often bleak conditions under which animals are kept before being slaughtered.

The last category of ignorance arising from the suppression of knowledge are taboos. While taboos can be actively cultivated, leading to similar responsibility issues than those to do with secrecy and denials, I will suggest that taboos within collectives could also be created unintendedly through the pressure to converge. Szanto (2017) describes how reciprocal irrational influences can reinforce themselves both top-down and bottom-up in small groups and organisational and corporate contexts alike. The group members feel the bond between them, and strive for unanimity and in-group cohesion. These reciprocal expectations can sometimes override rational assessment of the best course for action, leading to what Janis (1982) has termed "groupthink", where rationality and moral judgement deteriorates through in-group pressures.[9] We could imagine something like this taking place among the top managers at Exciting Electronics, especially if there has been no previous product recalls and there was emphasis on not making mistakes. Therefore the option of a large-scale product call had become a kind of a taboo at Exciting Electronics, not seriously even entertained by the top managers. Unintendedly created (i.e. non-cultivated) taboos mitigate blame to some degree at least, and perhaps in some cases they could even act as excusing condition. This does not, however, block forward-looking responsibility, as I will argue next.

[9] I thank Mikko Salmela for bringing the term to my attention. I find that groupthink could also be a contributing factor in an organisation resorting to denials.

4. Organisational design and the non-deliberate creation of ignorance

As I have been arguing, knowledge and ignorance are not evenly distributed within organizations. It is economical for an organization to consist of groups of experts that can work together when needed, as this allows for a wide range of skills and expertise to be employed in the organization. The process of specialization and coordination allows knowable recognised unknowns to be confined to or sustained within parts of the organization. When required, this ignorance can be overcome by bringing the different organizational actors together (Roberts 2013, 222). Ignorance, then, can be both necessary and even laudable, like with respecting privacy. However, organisational ignorance has problematic features also, as ignorance undermines the voluntariness and autonomy required for moral responsibility. In this section, I will argue that fragmentation of information in bureaucracies can lead to deliberate or non-deliberate creation of ignorance.

When we work together in collective settings, division of epistemic labour is not only very common, but also unavoidable. It allows organisations to absorb and process much greater amounts of facts than individual agents ever could. It also facilitates use of expert knowledge in overlapping areas and importantly the creation of new knowledge, ideas, inventions, and so on. We depend on others in our epistemic community for what Sandy Goldberg (2011, 121-122) calls *coverage* when we count on others to make relevant discoveries and reliably disseminate information of the same within the community.[10] As with secrecy, I argue that the division of epistemic labour and fragmentation of information in organisations is acceptable as long as the reasons given for it would be salient for objective outsiders.

While it is natural that all knowledge is not shared, in some cases it can be hard to draw a clear line on what information should be available to whom. In bureaucratic organisation there is a requirement for some ignorance for the focus to be on roles, rather than personal characteristics. Roles are

[10] Thank you to an anonymous reviewer for bringing this term to my attention.

narrowly defined for a more or less precise purpose that serves the organisation's goals, therefore ignorance of irrelevant personal characteristics of the people you deal with help things to run smoothly. An effective balance between informal relations and procedures, established in the course of frequent face-to-face contact, and the ignorance that is required for orderly procedures is thus required in any bureaucracy (Moore and Tumin 1949, 792-793). Here the ignorance is about facts that do not matter, so it is irrelevant and morally neutral.

However, organisations sometimes deprive individuals of their capacity to make good moral judgements by fragmenting available information. Recall the spy ring: maybe Shea would have refused to be part of the mission in the first place had she known details of the target and the method of the assignment. It could well be that the goal of the assignment goes against her values, and she was kept in the dark about the true nature of the mission on purpose to ensure her cooperation. To give a more humdrum example that is closer to real-life concerns, bureaucracy breaks work and knowledge into pieces, and bureaucratic compartmentalisation and the secrecy that often comes with it prevents information passing on from one department to another. This fragmentation of consciousness provides rationales for not knowing about problems, and for not trying to find out. Rational bureaucracy can, in this sense, stimulate irrationality (Jackall 1988, 194). Bureaucratisation is therefore never a purely technical matter, just a system of organisation, but a power system with privileges and domination. Max Weber already was worried about the implications of bureaucratisation for individuals' freedom and control, although he was supportive of bureaucracies as rational and efficient ways of humans to organise themselves.

Unlike Weber, Hannah Arendt (1970, 38-39) was very critical of bureaucracies and described them as "rule by Nobody". Bureaucracies can compartmentalise work to such a degree that individual human action is reduced to mere behaviour. If division of labour goes too far, people no longer know what their role is in the larger organisation, what their work is linked to, what the results are. Responsibility is impossible to locate anymore and becomes so diffused that the people working in the bureaucracy can come to view their actions to be outside the normal human realm where they would be responsible for what they do. Expanding on Arendt's thoughts, Larry May (1996, 71-76) similarly argues that organisational socialisation in bureaucracies can make people see themselves as the anonymous cogs of a machine, who do not have the need to develop a sense of responsibility in relation to what they do. Bureaucratic anonymity grows from the usual lack of face-to-face confrontation and not being directly linked to the consequences of one's actions. Some bureaucracies also socialise their members to feel that decisions should be made by the "experts" only, those members more experienced and knowledgeable. May (1996, 70) writes that

> bureaucratic institutions socialize people to see themselves not as actors but as those acted upon. The ensuing feelings of powerlessness can give rise to the acceptance of, and even participation in, harms these people [-] would never have found acceptable outside of the bureaucratic institution.

In addition to fragmentation of information, organisational frameworks also affect the way we think. Our minds both organise and censor our experiences through conceptual schemes. Werhane (1999, 85-95) describes how all of our activities are framed by mental models – our perspectives on things – and embedded in conceptual schemes. Our mental models are influenced by socialisation, culture, education, our upbringing, art, media, the place we work in. Our interests, desires, biases, intentions, and points of view operate as selective filters that restrict what we see in the world. Through the models, we make sense of our experiences, and interpret and clarify events to ourselves. This is often done retrospectively with events given a reframed focus and importance. We therefore do not observe the world objectively, but rather project our own perceptions on it and explain our experiences so that they fit our subjective point of view. We also tend to ignore data that does not fit our scheme. It is as if we are editing a movie and leave some of the scenes on the cutting room floor.

Thus corporate employees, for example, are trained to see things through the viewpoint of their employer, affecting the kinds of things they take into consideration when making decisions. The managers at Exciting Electronics could have been trained to think that they need to focus on introducing new product lines every quarter and keeping their prices as low as possible, for example. Maybe the ethos of Exciting Electronics gives low priority to proper quality control and safety measures in its supply chain, and places no emphasis whatsoever on the working conditions in its outsourced factories. If we choose any one perspective often, it gets reinforced in our minds. This is not to say that we have one-track minds, as most of us have several mental models to choose from so we can adapt to a given situation. Importantly, our perspectives can be altered if we choose to try to look at things from someone else's perspective. In any case, mental models could lead to viewing certain information as unimportant and outside the focus of the organisation.

When it comes to mental models, especially those actively cultivated by the organisation, it is easy to see how we could argue that certain organisational practices are for example negligent and should be changed. An example could be a corporation assigning no importance in its internal practices on looking at the working conditions at its suppliers, e.g. by leaving such considerations off the check-lists it has created for its brand managers. The situation with regards to blame is less clear with the other mechanisms that produce ignorance in organisational settings. I have been arguing in this section that fragmentation of information in bureaucracies can sometimes lead to non-deliberate creation of ignorance. When this has happened, and there is a harmful outcome, although not necessarily blameworthy for the outcome, the organisation should look at its design to try to make sure the same thing won't happen again. If they do not, I argue that they are blameworthy for being negligent.

5. Culpable ignorance

In the previous sections, I discussed how collectives can be culpable for the ignorance of their members. In this section, I turn to the possibility of culpable ignorance of individual members of the organisation (e.g. employees). In tracing cases of culpable ignorance (Smith 2016), an agent performs a morally inferior act from ignorance that can be traced back to an earlier act that created the conditions for ignorance. While this act of unwitting misconduct is excused by ignorance, the agent's earlier act is not, so the individual might be blameworthy. I will argue in this section that the earlier act that creates the conditions for ignorance can take place at the collective level, so blame should lie there also.

Assigning culpability to an agent for her ignorance is not yet to assign blameworthiness. Holly M. Smith (1983, 552) argues that all cases of culpable ignorance involve a sequence of acts: the initial act (the "benighting act") where the agent fails to improve their cognitive position, resulting in ignorance, and a subsequent act where the wrongful act is done due to this culpable ignorance. Smith further observes that frequently the benighting act takes the form of an omission, like failing to learn or find out something. The benighting act affects which subsequent acts are available to the agent, leading to the optimum act not being either epistemically or physically available to her.[11] According to Smith (2016), while the agent is not blameworthy for the act that was done in culpable ignorance, they are to blame for the earlier failure to obtain the information that would have led to her not being ignorant in the relevant manner. The agent has performed an act that is morally inferior to the counterfactual act she would have performed had she obtained all the necessary information. The ignorance is thus traceable to a past epistemic negligence.

To return to our example once again, it seems clear that Violet in not to blame in the tale of exploding toasters, as there is no way she could have known what was going to happen. She relied on what she heard from Exciting Electronics and in this way Violet was *epistemically dependent* on what Sharon told her about the products.[12] The case is not so clear with Sharon. Smith (1983, 544-547) presents three types of cases where ignorance does not excuse, as the person should have realized what they were doing. *Deficient investigation* is the first type, either through failing to investigate properly, or failing to investigate at all. While this is not the case with Violet, Sharon could fall into this category if it were the case that

[11] Alternative take on culpable ignorance is that culpability arises from holding of beliefs (Sher 2009).

[12] See Goldberg 2011 for a discussion on *epistemic dependence.*

she had failed to read an internal memo about possible problems with the new manufacturers' products. *Preventing subsequent discovery* presents the second case: a person has either failed to remove or introduced a condition which made it impossible for him to acquire true belief of *x*'s wrongness. If information about faulty products was available in print at Exciting Electronics but Sharon missed this because she has never learnt how to read properly (a life-long secret she has just about managed to hide from her employees), she would fall into this category. Finally, culpable ignorance could arise from *deficient inference*: had the agent made the inference warranted by his background beliefs, he would have correctly believed the act to be wrong. To use Sharon once again, had she remembered that her colleague Bob told her about problems in some new factory, she would have put two and two together when she heard from her manager Bianca that Exciting Electronics has changed its toaster manufacturer.

Although Smith's influential work on culpable ignorance looks at cases of individual responsibility, I see no reason why it cannot be framed in organisational setting, allowing for the benighting act to be done by a different person than the morally inferior act that follows. Sharon did not disclose information to Violet about the faults, so Violet ordered faulty toasters to all shops in the retail chain. Had she known about the loose wires, she would have not completed the purchase. Violet's ignorance is not culpable here, though, while Sharon's might be. I gave examples earlier how Sharon's ignorance could be traced to either deficient investigation, deficient inference or preventing subsequent discovery. It could also be the case that Sharon's and Bianca's ignorance is not culpable either, but the result of suppression of information by Exciting Electronics, or a knowable recognised unknown for the organisation.

Compartmentalisation of information raises the further possibility that ignorance is produced systematically at Exciting Electronics, without it tracing back to the action or omission of any one person, or even a group of people. It could be several acts by several agents within the organisational setting that together produce the ignorance. Indeed, the deficient inference could take place at the collective level, with the organisation failing to make the inference warranted by their background knowledge and beliefs due to fragmentation of information, for example. In this way, the benighting act happens at the organisational level and any possible blame should be directed at the collective.

In an organisational setting the benighting act can either be done by an individual or take a collective form, in which case it is harder to point out exactly whose failure it was. Blameworthiness in a collective setting is a complex concept, so it will be harder to give simple or general answers. Depending on how they are used and how justified their usage is in the first place, denials, secrecy, and taboos do not excuse everyone in an organisation, as they are usually instruments of power. They fall under suppression of knowledge at the top managerial level, or whatever level engages in the behaviour, while they can result in either excusing ignorance at the bottom level, or lead to conditions where it is all too easy to fall into the trap of making deficient inferences. If the parameters of a given role dictate acts that lead to culpable ignorance in others, the moral responsibility for that ignorance falls (either fully or at least in part) to the collective. The requirements of roles and an individual's leeway within them is a fascinating area for moral responsibility, but it falls outside the scope of this paper.

As I stated above, the benighting act can also take a collective form. I stipulated that the loose wires are so because of the too hectic manufacturing process. If competitiveness of the prices of goods takes precedent over all other considerations in their ethos, Exciting Electronics is taking a risk that falls within known risks. They might have been very lucky in the past and gotten away without adequate safety checks, but they could not justify this way of operating by appealing to ignorance: the knowledge would have been found out given the right resources and motivation. This would be an example of deficient investigation and the resulting ignorance is blameworthy.

Regardless, there is a potentially interesting consequence for work on individual culpable ignorance. I have argued that in an organisational context the benighting act, i.e. the earlier act that creates the conditions for ignorance, can take place at the collective level, so blame lies there also. If my argument works, then the quality of will of an agent is not necessary for blame, although it remains sufficient. The reason for this is that in an organisational setting there needs to be no bad will: organisational fragmentation of information alone can produce morally inferior acts. Bad will in this context should be understood as morally objectionable aversions or desires, or lack of proper moral concern. It often grounds the blameworthiness of agents in culpable ignorance literature. This common assumption is shared by Smith, who includes it in her account of moral blameworthiness.[13] To give two other recent examples, Jan Willem Wieland (2017) suggests that individuals are blameworthy for their strategic ignorance depending on their moral concern, while Gunnar Björnsson (2017) conceptualises quality of will of an agent through caring enough about how well things go and argues that ignorance fails to excuse when someone should have cared more.[14]

In an organisational setting there needs to be no bad will in order for there to be culpable ignorance. Violet and Sharon (and the workers in the factory where the toasters are put together) all lacked a morally objectionable configuration of aversions or desires. Sharon did not mislead Violet because she wanted to cause danger around the breakfast tables of Sell-A-Lot's toast-loving customers, or because she did not care enough about customer safety. It could very well be that she unwittingly lied about the quality and safety of the product because the relevant information was too fragmented at Exciting Electronics: division of labour between departments was too deep, lines of communication were unintentionally complicated, and so forth. Sharon relied on the epistemic coverage of her colleagues, but was let down in this regard. The act of telling the customers something that was not true was unwitting, as she would have acted otherwise had she been privy to the information about the faulty wires. However, from Violet's point of view Sharon's sale pitch about the

[13] Smith 2016, 98: "I shall employ a 'quality of will' account according to which it is the quality of the agent's motivations in performing the blameworthy act that make her worthy of condemnation for performing it."

[14] The assumption is of course widely shared within moral responsibility literature, the quality of will account of blameworthiness is not exclusive to culpable ignorance literature.

high-quality of their toasters was a lie, taken as a statement representing the knowledge of Exciting Electronics.

Importantly, if the organisational design was unintentionally so that it caused fragmented information, there is no bad will at the level of the ethos of the organisation either. The benighting act takes a collective form, so the culpability is that of the collective also. We could of course come up with a version where the bad will of a manager in the factory that supplies Exciting Electronics with its machines, or a manager at Exciting Electronics or Sell-A-Lot, is the cause for the resulting organisational ignorance. This could be either due to personal failings on the part of the manager, or due to being too influenced by certain harmful organisational mental models. But this is beside the point: bad will is not necessary for the examples to get off the ground. Organisational fragmentation of information alone can produce morally inferior acts. Although there is no bad will, if the organisation does not take action to try to change its internal practices and communication flows after it has been made aware of the problems, then we can argue that it should care more. After all, it has come to know that the knowledge it has about the safety of its outsourced products is not robust enough to prevent such large-scale errors from happening. Still, some organisational fragmentation of information is always necessary and can lead to unintended consequences.

More generally, all collective action can result in outcomes that were completely unintended. Smith argues that there is a degree to which we can be held accountable for the consequences of our actions, and this is linked to the outcomes falling within the predictable outcomes for that act. In other words, culpable ignorance arises only when the unwitting wrongful act falls within the known risks of the earlier act that infects the later act (Smith 1983, 551). This makes a separation between knowingly risking something and having no reason to believe that the benighting act would result in a wrongful act. Christopher Kutz's (2000) approach is somewhat different: while he acknowledges that our accountability for consequences that flow from our actions is in theory infinite and therefore needs to be "normatively delimited", he notes that some response is warranted even when the outcome is unforeseeable (pp. 142-143):

> By taking responsibility for the consequences of our acts, we demonstrate to others a concern for their projects and interests, and thereby work to ensure their respect for our work. Within this delimited set of consequences, normative questions of individual response arise: whether to apologize, compensate, or repair.

Kutz illustrates this point with a wasp that enters the house when he lets the cat out. He did not intend to let the wasp in, but when it stings you badly, he should express sadness at your pain and offer you comfort. He is not at fault and resentment towards him is therefore not warranted. The response to the pain caused by the sting indicates the importance attached to your interests, and any claims for him to respond are rooted in the fact that his agency led to the suffering, however unintended it was. In other words, "accountability for unintended consequences manifest an acknowledgment of the fact that one's projects have interfered with another's interests" (p. 143).

This applies also to the unintended consequences of collective action, as we are complicit in the consequences of what we do together. If some harm is a direct consequence of what we do intentionally, even though the harm itself was not intended, we have a duty to acknowledge it in the appropriate way (i.e. apologise, compensate, etc.). If no apology or compensation was forthcoming from Sell-A-Lot to their customers, although the fault was with Exciting Electronics, the customers of the retail chain would be right to blame the company for not caring for their customers sufficiently.

But does it make sense to apply an account of reactive attitudes towards organisations in the first place? What, exactly, are we blaming when we are blaming a corporation, for example? Imagine an error done by a wholly automated organisation.[15] Would liability work better here than assigning blame?

There are a number of possible routes to take to answer this question. One could point out that we do in fact blame collectives in our everyday lives and that it makes sense to blame the corporation because that is what we do. Reactive attitudes could be thought to be directed at the relevant

[15] I thank Pekka Väyrynen for suggesting this option.

members of the corporation, the ones who could make a difference, or towards the ethos. Or one could argue that reactive attitude target the collective itself and that we should understand them in functional terms (like Hess and Björnsson 2017 do). What I want to argue here is that automation itself does not block out reactive attitudes.

Let us say that Exciting Electronics is ahead of its time, and is using an algorithm to make decisions about outsourcing its production. A complicated programme calculates the most cost-efficient supplier, with the best price-quality ratio, and sends out the necessary paperwork. The toasters start exploding. There was no option here for deficient inference in the way I described earlier with Sharon, Bob and Bianca: the algorithm alone knew about the switch in suppliers. Once it was informed about problems in the factory by the supplier (via an internet form), it calculated these to fall within acceptable risk parameters. Importantly, those parameters have been set by somebody. They are akin to engineering decisions that go into making automated vehicles. While the options might be set by engineers, the parameters of acceptable risks are decided by the corporation. Again, the ethos of Exciting Electronics could be argued to include negligence and to be blameworthy.

6. Concluding remarks

This paper has discussed how ignorance in an organisational setting is a complex phenomenon, and how it can be neutral, praiseworthy or blameworthy. Ignorance is a necessary ingredient to any organisation and it serves many purposes, from safeguarding trade secrets and our right to privacy, to ensuring the smooth running of daily operations. Ignorance is also a powerful tool for organisations to influence their members and stakeholders.

I have argued that the ethos of an organisation can be blamed for being morally lacking in some way, for example allowing negligence towards the safety of their customers or workers in their supply chain. While ignorance can be produced knowingly, it can also be an unintended side-effect of bureaucratization. If problems emerge, or a likelihood of bad outcomes is pointed out to them, the organisation has a forward-looking duty to try to fix its design. If the organisation fails to respond, they are blameworthy for failing to improve their design or organisational practices when it comes to information flows.

One thing is certain: we should not ignore ignorance and its many facets when attempting to analyse organisations and other collective phenomena.

Acknowledgements

This paper has benefited from comments by the two anonymous reviewers and audience members at Philosophical Society of Finland Colloquium 2018. I would also like to thank the audiences at Collective Intentionality IX conference at Indiana University, Bloomington, and the subsequent invited talk at Marquette University, the 2014 Congress for Doctoral Students in Philosophy at University of Tampere, and the 2016 Stockholm-Helsinki graduate workshop in Practical Philosophy at Stockholm University for their very helpful comments on earlier versions of the argument. Last but not least, I would like to thank Arto Laitinen and Pekka Mäkelä for written comments on an early draft. The research was made possible by two personal grants from the Finnish Cultural Foundation, for which I am grateful.

University of Helsinki

References

Arendt, Hannah (1970), *On Violence*. Harcourt Brace & Company, San Diego.

Björnsson, Gunnar (2017), "Explaining (Away) the Epistemic Condition on Moral Responsibility", in Robichaud and Wieland (eds.): *Responsibility: The Epistemic Condition*. Oxford University Press, Oxford, pp. 146–162.

Björnsson, Gunnar and Kendy Hess (2017), "Corporate Crocodile Tears?: On the Reactive Attitudes of Corporations", *Philosophy and Phenomenological Research* 94(2), pp. 273–298.

Dufresne, Ronald L. and Evan H. Offstein (2008), "On the Virtues of Secrecy in Organizations", *Journal of Management Inquiry* 17(2), pp. 102–106.

Galison, Peter (2008), "Removing Knowledge: The Logic of Modern Censorship", in Proctor and Schiebinger (eds.) *Agnotology: The Making and Unmaking of Ignorance*. Stanford University Press, Stanford, pp. 37–54.

Goldberg, Sandy (2011), "The Division of Epistemic Labor", *Episteme* 8(1), pp. 112–125.

Hormio, Säde (2017), "Can Corporations Have (Moral) Responsibility Regarding Climate Change Mitigation?", *Ethics, Policy & Environment* 20(3), pp. 314–332.

Isaacs, Tracy (2011), *Moral Responsibility in Collective Contexts*. Oxford University Press, New York.

Jackall, Robert (1988), *Moral Mazes: The World of Corporate Managers*. Oxford University Press, Oxford.

Janis, Irving (1982), *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Wadsworth, Boston.

Kutz, Christopher (2000), *Complicity: Ethics and Law for a Collective Age*. Cambridge University Press, Cambridge.

Laitinen, Arto (2014), "Collective Intentionality and Recognition from Others", in Konzelmann Ziv and Schmid (eds) *Institutions, Emotions, and Group Agents: Contributions to Social Ontology*. Springer Science+Business Media Dordrecht, pp. 213–227.

Lu, Lauren Xiaoyuan and Jayashankar M. Swaminathan (2015), "Supply Chain Management", *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, pp. 709–713, doi.org/10.1016/B978-0-08-097086-8.73032-7

May, Larry (1996), *The Socially Responsive Self: Social Theory and Professional Ethics.* The University Chicago Press, Chicago.

Moore, Wilbert E. and Melvin M. Tumin (1949), "Some Social Functions of Ignorance", *American Sociological Review* 14(6), pp. 787–795.

Peels, Rik (ed) (2016), *Perspectives on Ignorance from Moral and Social Philosophy*. Routledge, London.

Proctor, Robert N. and Londa Schiebinger (eds.) (2008), *Agnotology: The Making and Unmaking of Ignorance.* Stanford University Press, Stanford.

Roberts, Joanne (2013), "Organizational ignorance: Towards a managerial perspective on the unknown", *Management Learning* 44(3), pp. 215–236.

Sher, George (2009), *Who Knew? Responsibility Without Awareness*. Oxford University Press, Oxford.

Smith, Holly (1983), "Culpable Ignorance", *The Philosophical Review* 92(4), pp. 543–571.

Smith, Holly (2016), "Tracing Cases of Culpable Ignorance", in Peels (ed.) *Perspectives on Ignorance from Moral and Social Philosophy*. Routledge, London, pp. 95–119.

Szanto, Thomas (2017), "Collaborative Irrationality, Akrasia, and Groupthink: Social Disruptions of Emotion Regulation", *Frontiers in Psychology* 7: 2002.

Tuomela, Raimo (2007), *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press, New York.

Werhane, Patricia H. (1999), "The Very Idea of a Conceptual Scheme", in Donaldson, Thomas; Werhane, Patricia H. & Cording, Margaret (eds) (2002): *Ethical Issues in Business: A Philosophical Approach (7th Edition)*. Pearson Education, New Jersey, pp. 83–97.

Wieland, Jan Willem (2007), "Responsibility for strategic ignorance", *Synthese* 194: pp. 4477–4497.

Davidson on Aristotle and Philosophy of Action[1]

HARRY ALANEN

For some, Aristotle may appear to have developed a causal theory of action of the kinds later developed by Donald Davidson and others. This is understandable given that Aristotle repeatedly claims that desire is the cause or origin of action and of animals moving themselves with respect to place.[2] Indeed, Davidson himself remarks that "Aristotle pretty much invented the subject [of action] *as we now think of it*."[3] The aim of this paper is to show that Davidson is mistaken regarding this claim. This is significant because unless we are careful about what current assumptions govern our ways of thinking about philosophy of action, we may impose those assumptions on our predecessors. In so doing, we risk making our history anachronistic, and blinding ourselves to unique ways of approaching questions about action and agency. Despite the superficial similarity between Aristotle's views and contemporary causal theories of action, his arguments are made within a very different historical context and philosophical framework. Hence his views on action differ from contemporary causal theories.

[1] Previous versions of this paper have been given at Oxford, Helsinki, and Yale – I thank the audiences at these occasions for their questions and comments. I also wish to thank Lilli Alanen, Deborah Brown, Ursula Coope, Michael Della Rocca, and two anonymous reviewers, whose suggestions helped clarify my arguments.

[2] See for instance *De Motu Animalium* (MA) 6 700b15-25, 7 701a33-36, *De Anima* (DA) III.7 431a8-16, III.9-10, *Ethica Nicomachea* (EN) III.1 1111a22-b3, EN VI.2 (= *Ethica Eudemia* (EE) V.2).

[3] Davidson 2005, 277, emphasis added.

While developing Aristotle's philosophy of action would be a worthwhile project, my aim here is more general and preliminary. In order to do Aristotle's views and his historical context justice, one must first be clear on what kinds of questions and problems he is trying to answer. This is true for doing any history of philosophy. This means one cannot uncritically adopt contemporary ways of doing philosophy, nor raise the same questions that are currently of interest, without affecting one's interpretation. Davidson (and others) get Aristotle's views on action wrong because they do not sufficiently consider that philosophy of action might concern itself with quite different questions than those proponents of causal theories typically assume. My aim here is to challenge some of those assumptions. I begin by sketching out Davidson's views on what kinds of issues a serious philosophy of action should concern itself with (§§1-2). Davidson is significant since his approach to philosophy of action has been widely influential (and continues to be so). I then proceed to discuss his interpretation of Aristotle and why Aristotle may appear as a proponent of a causal theory of action (§3). I then show that Davidson's approach to philosophy of action is itself a product of certain important developments in the history of philosophy. These developments are themselves rejections of different kinds of *Aristotelian* positions or views (§4). In various ways, these developments set the agenda for how questions about life, action, and emotions are raised and answered in the 20th century – the century during which philosophy of action becomes a distinct field in its own right.[4] This should raise doubts that Aristotle's approach to philosophy of action is similar to the one Davidson's envisions.

By showing how certain developments in the history of philosophy affect contemporary philosophy of action – causal theories of action in particular – I hope to do two things: to show that one should be wary of appealing to historical authorities as the representatives of contemporary views, and to invite further discussion about the history of philosophy of action, a topic that remains underexplored. Getting clear on the questions that guide our thinking, and the assumptions

[4] For an overview of the development of philosophy of action as a field, see Stoutland 1989, and 2011a.

these rest on, allows us to better evaluate whether or not our philosophical predecessors offer answers to the problems that concern us today. This gives us a better understanding of historical views. History of philosophy and contemporary philosophy are linked, not because philosophers have always dealt with the same issues, but because researching the history of philosophy is a way of doing contemporary philosophy. If I can show that Davidson's interpretation of Aristotle is mistaken because it rests on assumptions about what philosophy of action is about – assumptions Aristotle need not share – then this should clear the way for a more careful treatment of Aristotle's views about action and agency. At the same time, this should pave the way for approaching philosophy of action in the history of philosophy from a more nuanced standpoint.

§1. What Kinds of Questions does Philosophy of Action seek to Answer?

As we saw above, Davidson credits Aristotle with inventing "the subject as we now think of it". There are two questions we should raise at this point. The first is: what is "the subject" Davidson has in mind? The other: what is the way in which "we now think" of this subject, *according to Davidson*?

The answer to the first question is, naturally, "philosophy of action" – but what does this mean? Is philosophy of action primarily concerned with questions about *what actions are*, or, perhaps, about *what it is to be an agent*? Or both? Perhaps there is no substantial difference between these two ways of construing the fundamental question of the subject?[5] The second question in turn is about how we answer the questions that are raised in our subject-matter. The questions are connected: depending on what we take as our question(s), and depending on the ways we answer these questions, our understanding of what philosophy of action is, and how it is to be done, will differ.

Let me try and clarify this further. There are a number of different ways one might understand such questions as

[5] For the sake of simplicity, I will restrict myself to these two options as the fundamental questions for philosophy of action.

"what are actions?" or "what is it to be an agent?" For example, they might be taken as metaphysical questions: what sets actions or agents apart from other entities one might postulate in one's ontology (such as events and states, or non-living substances and mathematical entities)? Answers to metaphysical questions tell us something about the world, about what it contains. But the questions might also be understood in other ways. For example, they might be conceptual questions: questions about what we mean when we speak of "actions" or "agents", or what the use of such concepts entails. Or they might be understood as psychological questions, in which case one's project would be to say what is psychologically distinctive or true about intentional actions, or about agents when they behave in certain ways. The differences between these ways of understanding the questions need not be sharp. Rather, they will depend on further views one may have about how the world, language, and psychology relate to one another.[6] This point is worth keeping in mind. We should try and be as clear as possible when articulating what we take as our fundamental question(s), and what kind of a question we are articulating (e.g. if it is a metaphysical question or not), and how it relates to other questions and topics (are actions explained in terms of the agent's psychology, and do such explanations derive their truth by corresponding to observable mental states in the agent)? It is important to make such assumptions explicit because they determine what we will consider as acceptable answers to our questions. In particular, care needs to be taken when we begin to investigate historical views in philosophy. We must be careful not to import our present-day assumptions to our interpretation of historical views (without critical investigation of them first).

§2. Davidson and "Davidsonian" Philosophy of Action

One thing we must thus be clear about is what kind of a question we are raising when asking what actions are or what it is to be an agent. Another is what relation there is between an action and an agent, and whether either concept is more fun-

[6] For example, our answers to conceptual questions may have metaphysical implications. Thanks to Michael Della Rocca for helping me emphasize this point.

damental than the other, or if both must be given equal consideration in a theory of action. For example, one might seek to elucidate what actions are by taking agency as basic or more fundamental, or one could seek to explain agency by determining what things in the world are actions. Davidson favours this latter approach and spelling out his views on this question should help us see the distinctive features of his philosophy of action, and more generally, of the so-called "standard story" of action which Davidson helped inspire.[7]

In "Agency", Davidson begins by asking "What events in the life of a person reveal agency; what are his deeds and his doings in contrast to mere happenings in his history; what is the mark that distinguishes his actions?"[8] The implicit thought here is that the concept of agency is contained within the concept of events, such that if we successfully distinguish actions from *other events* (the sc. "mere happenings"), we will also distinguish the episodes of the person's life where they are agents (and not patients, who only suffer or undergo what happens to them). This approach assumes that actions are events, or that we need a notion of events to fully understand what actions are.[9]

Davidson's answer to his initial question is complex. According to him "a person is the agent of an event if and only if there is a description of what he did that makes true a sentence that says he did it intentionally."[10] Both the notion of "event" and of "descriptions" are central here, and I cannot hope to do full justice to the details of his views. For Davidson one and the same event can have many different descriptions, and to describe an event as an action is to describe that event as an agent's intentional doing. Actions can be described both in terms of the intended features or consequences, and in terms of unintended ones. Although Davidson does not attempt to give a reductive account of agency, he nonetheless thinks we can help clarify the notion

[7] There are important differences between the kind of causal account of action Davidson developed, and the so-called "standard" causal story; I chart some key differences below.

[8] Davidson 2001, 43.

[9] For a similar point, see Hornsby 2004, 4n5.

[10] Davidson 2001, 46.

of agency (to a certain point) by introducing the notion of causation. To show that an agent is the cause of an event is a way to justify the initial attribution of agency.[11] Causation is a relation that holds between events, and to say that an agent is the cause of an event is an "elliptical" way of saying that something the agent did, i.e. an event described as an action, is the cause of some further event.[12] Causal explanations in terms of events derive their explanatory power from the fact that we can, in principle, cite laws of nature that connect the two events as cause to effect.[13] However, Davidson holds that event causation cannot fully be employed to explain the basic sense of agency – it cannot properly explain the relation between an agent and her action:

> [...] event causality cannot [...] be used to explain the relation between an agent and a primitive action. Event causality can spread responsibility for an action to the consequences of the action, but it cannot help explicate the first attribution of agency on which the rest depend. (Davidson 2001, 49)

However, in a footnote to this passage Davidson adds that agency can be analysed "in part" in terms of event causation.[14] By "partial" Davidson means that we cannot hope to give a reductive account of agency, where agency is explained in terms of one event (such as a desire) causing another (the primitive action). [15] By contrast, a "full" analysis or explanation of agency, would require specifying laws governing the relation between the agent's reasons – her beliefs and desires – and the events described as her actions, something Davidson viewed as not possible since he denied the existence of strict, psychophysical laws, which would govern those relations.[16] But without describing *some* event as the

[11] Cf. *ibid.*, 48.

[12] Cf. *ibid.*, 49.

[13] Cf. *ibid.*, 52-53.

[14] *Ibid.*, 49n7.

[15] By "primitive actions" Davidson means an action that an agent does without doing anything else. For example, I cook a risotto *by* moving my body in certain ways, but I do not (ordinarily) move my body *by* doing anything else.

[16] In the initial publication of "Actions, Reasons, and Causes" Davidson suggests we could give the necessary and sufficient conditions for inten-

primitive action of the agent, we cannot hope to analyse other events, such as the consequences of her primitive actions; which is how we can attribute agency to someone, as noted above. However, to describe some event as the agent's primitive action does mean we can analyse it in terms of the agent's beliefs and desires, which are causal concepts. While Davidson holds that mental states are not themselves events, coming to have them entail events, and hence the concept of agency is "partially open" to an event-causal analysis.[17] An attribution of agency might be basic, but it nonetheless involves events. Thus, even if we cannot reductively explain agency in terms of events, agency is still *revealed* by events described as an agent's primitive action.

On Davidson's approach to agency, event causation plays an important role since it helps justify the attribution of agency by showing that a putative action of the agent (an event) has a further consequence (another event), that is, we use event causation to explain the relation between an agent's primitive action and other actions or events the agent causes by moving her body. Finally, event causation can be used in a partial analysis of the concept of intention or agency. Although Davidson does not think a reductive account of agency (and thus, of action) is possible, we nonetheless refer to our reasons such as our beliefs and desires as the causes of actions, and thus we make use of causal concepts ("belief" and "desire"), which entail the existence of events. This entailment allows Davidson to say that our reasons *cause* actions, since both the actions and the reasons that cause them either are, or entail, events, and given his view that the causal relata are events.

tional action, but he changed his mind, coming to think that we cannot specify the sufficient conditions – compare Davidson 1963, 693n5 with Davidson 2001, 12n5; a reason for this change of mind was that Davidson accepted the force of the problem of deviant causal chains (cf. Davidson 2001, 79-80, and Davidson 2001, xvii).

[17] For Davidson, a full analysis of agency would presumably entail specifying the events that stand in the causal relation, and the laws that connect these events. However, Davidson does not think we can give such explanations, since – for him – there are no strict psycho-physical laws that would allow us to predict or explain mental events; this is the "principle of the anomalism of the mental".

Given the importance of events in Davidson's approach, we will want to know something about what Davidson takes events to be, and his views on causation, and explanation. We are already familiar with his claim that events allow for different ways of describing (and thus referring to) them. But what are events? According to Davidson, events are unrepeatable particulars or "concrete individuals",[18] i.e., each individual event has a distinct time and place.[19] If one event is the cause of another, this means the cause-event must be temporally prior to the event which is the effect. Further, if one event causes another, then there is a law of nature that connects them.[20] In these respects Davidson's account of causation is Humean in character.[21] While we need not know what those laws are, causality entails the existence of strict laws.

However, how can we determine if one event *is* the cause of another? Causality is a relation that holds only between events, and although one event can be described in many ways, a causal relation holds no matter how they are described. Thus, not all descriptions of events are explanatory. As Stoutland helpfully puts it:

> Ascriptions of causal relations need not, therefore, *explain* phenomena: Saying truly that what Karl referred to last night was the event-cause of what happened to Linda a year ago does not explain what happened to Linda a year ago.[22]

Sentences describing events are *extensional*, whereas explanations are *intensional*, since

> to explain phenomena is always to explain them *as* such and such, that is under a description [...]. The point of an explanation is to render phenomena intelligible, and what does so under one description of the phenomena may not do so under another.[23]

[18] Cf. Davidson 2001, 181

[19] Cf. Davidson 2001, 309-310.

[20] Cf. Davidson 2001, 208; see also Essay 7.

[21] For an overview of differences between Davidson and Hume, see Ehring 2014, 286-7.

[22] Stoutland 2011b, 298.

[23] *Ibid.*, 298-299. Cf. Davidson 2004, 110.

There is a thus an important difference for Davidson between causal relations and explanations, and not all true descriptions of events are explanatory.

One way to explain a phenomenon is by citing the relevant law of nature, but since Davidson does not think there are strict laws governing human actions, such events cannot be explained by appealing to laws. Instead, actions are explained in terms of the agent's primary reasons – her beliefs and desires – and the explanations in terms of these Davidson called "rationalizations".[24] Rationalizations, or reasons-explanations, count as causal explanations because beliefs and desires are causal concepts as they entail there being events, they are cited to explain other events (actions and their consequences), and because reasons make the action-events intelligible. Although there is no strict law explaining that someone who believes and desires something will invariably do something, knowing an agent's primary reasons allows one to say what someone would generally do, or what they tend to do, all things being equal. Since such generalizations require *ceteris paribus* conditions, they cannot be taken as strict laws.[25] According to Davidson "a belief and a desire explain an action only if the contents of the belief and desire entail that there is something desirable about the action, given the description under which the action is being explained. This entailment marks a normative element, a primitive aspect of rationality."[26] This normative element is an irreducible feature of action-explanations, because in order to explain an action we must first begin by identifying or describing some event as an action, and because "reason-explanations make others intelligible to us only to the extent that we can recognize something like our own reasoning powers at work."[27] This does not mean that actions or other mental concepts cannot be studied systematically; only that

[24] See Davidson 2001, Essay 1.

[25] Cf. Stoutland 2011b, 300.

[26] Davidson 2004, 115.

[27] Davidson 2004, 114-115.

any such systematic approach will be "a science of rationality" – and not a physical science.[28]

Let me make one further point about events in Davidson's philosophy, and their relation to descriptions. According to Davidson:

> We constantly identify or describe things partly in terms of the causal relations. [...] Many verbs incorporate this idea: if John breaks a window, something he did caused a window to break. [...] Causality is a relation between events. So *to grasp what it means to say that John broke the window, we need to invoke the existence of two events*: we are saying, "There were two events: one was something that John did, one was a breaking of the window, and the first event caused the second." [...] If there are two events, their times may be different, and if one caused the other, the time of the first must be before the time of the other. So "John threw a stone" and "John broke the window" can involve just one *action*, but two events, because "John broke the window" just means John did something (in this case threw a stone) which caused the window to break.[29]

On this approach simple sentences like "John broke the window" entail there being two events – despite the sentence only referring to John, the window, and the breaking of it. This is because the sentence means that there are two events, one of which is John's action that causes the window to break. This allows Davidson to explain the relation between transitive and intransitive forms of verbs, in this case: "the transitive 'break' means 'cause to break'."[30]

What the preceding discussions show is how Davidson's anomalous monism, and his views on causation and explanation are vital for a correct understanding of his philosophy of action. Although few agree with Davidson's views on all these topics, his way of doing philosophy of action has nonetheless been highly influential on other causal theories of action. In particular, his views on events and causation have inspired more straight-forward causal theories such as the sc. "standard story" of action. Before looking at how Davidson

[28] See Davidson 2005, 291; cf. Davidson 2004.

[29] Davidson 2005, 287, emphasis added.

[30] *Ibid.*; cf. Davidson 2004, 104-105.

may have inspired later causal theories, it will be helpful to summarize the main features of Davidson's approach to action. In this way the differences and similarities between the two can be more clearly articulated.[31]

As we've seen, Davidson takes the nature of action as a central question for philosophy of action. If we can determine what an action is then it will also become apparent when someone is an agent. In this sense, questions about the nature of action – a metaphysical question – are prioritized over questions about agency (and responsibility). While Davidson does not think we can give a reductive account of action or agency, in explaining action we describe and interpret certain events as actions, and we explain actions by citing an agent's primary reasons. While the explanation of action thus has irreducible normative feature, reasons explanations count as causal explanations since beliefs and desires are causal concepts. Although beliefs and desires entail events, he holds that we cannot formulate any strict psychophysical laws which would allow us to predict behaviour caused by them. Finally, in describing an event as an action, we either describe it as a cause of another event, or in terms of what the action caused, since verbs that denote action entail that there are two events, the action being the cause-event of another event.

Obviously not everyone agrees with Davidson's conception of events, or his anomalous monism.[32] However, his claims that beliefs and desire cause and explain actions have been developed to what is now often called the standard story of action.[33] According to this story actions are bodily movements caused (in the right kind of way) by an agent's beliefs and desires (or other mental events of the agent, or neurophysical events in the agent's brain). Often this is given as a more straightforward casual theory than Davidson's. Recall that Davidson treats causal relations and causal explanations independently. On a more straightforward view, causes explain phenomena because causal explanations refer to the

[31] For the sake of brevity, I'll focus on philosophers working on causal theories of action within an events-based framework.

[32] For an overview of different ways of thinking about events see e.g. Casati & Varzi 2015.

[33] The differences between Davidson and the standard story are clearly laid out in Stoutland 2011b.

events which function as causes. Applied to a physicalistic ontology, actions are bodily movements caused by agent's believing and desiring; such an approach enables one, in principle, to seek a reductive account of action by explaining actions in terms of the events that cause it.[34] One point of disagreement between Davidson and others is thus whether or not one can give reductive accounts of intentional action (and thus agency). For example, Michael Smith has suggested that Davidson's basic picture must be "supplemented" in order to give us a reductive account.[35] If appropriately developed, Michael Smith argues, the standard story "purports to tell us what makes someone an agent, rather than a mere patient."[36] This difference is made clear by spelling out "the causal etiology of what happens when a body moves."[37] According to Smith a bodily movement counts as an action only if it is caused in the right kind of way by the agent's beliefs and desires.[38] By contrast, for Davidson an action is an event described appropriately, whereas proponents of the standard story typically identify the agent's bodily movement as the agent's (basic) action.

Despite such disagreements, proponents of event-causal accounts of action typically hold that the nature of actions is best determined by examining the causal relations that obtain between events: on this view, actions are events that have been caused in a particular way, by the events that are the agent's beliefs and desires. So despite certain disagreements in the details, there is a general commitment to approaching philosophy of action in a rather distinctive way, by focusing on certain metaphysical questions (such as the nature of action and causation), and by developing a philosophy of mind according to which psychological phenomena such as beliefs

[34] One issue with this approach is that it seems to leave the person no role as an agent; she is mere a container in which certain events take place which cause her actions.

[35] Cf. Smith 2004, 167; Smith 2012, 397-400.

[36] Smith 2012, 400.

[37] *Ibid.*, 387.

[38] For a discussion of various ways of trying to avoid the problem of causal deviance (and the problems these attempts face) see Mayr 2011, Chapter 5.

and desires explain and/or cause actions or bodily movements.[39]

We should now turn to Davidson's interpretation of Aristotle to see to what extent his interpretation matches his own approach to action.

§3. Davidson on Aristotle on Action

Aristotle's repeated claims that desire (*orexis*) is the cause or origin (the *aition* or *archê*) of action (*praxis*, *prattein*) and of animals moving themselves with respect to place (*kinêsis kata topon*), may appear to some as an early formulation of a causal theory of action similar to ones developed by Davidson, and others.

For example, in *On the Movement of Animals* Aristotle writes that: "This, then, is the way in which animals are impelled to move (*kineîsthai*) and act (*prattein*): the proximate reason for movement is desire (*orexeôs*), and this comes to be either through sense-perception or through *phantasia* and thought."[40] The role of desire in animal locomotion is further discussed in *DA* III.9-10, including its relation to practical thought (in animals capable of both desire and reason): "Hence there is good reason for the view that these two are the causes of motion, desire and practical thought. For it is the object of desire which causes motion; and the reason why thought causes motion is that the object of desire is the starting point of thought."[41] In the sc. "Common Books" of the *Ethics*, desire is described as an origin of decision (*prohairesis*), and thus of action and production (*poiêsis*) and the movements involved in these.[42] Further, in EN III.1 Aristotle claims that the origin of an action that is voluntary (*hekousion*) is in oneself (*en autoî*), going on to suggest that actions done because of certain kinds of desires should also be counted as voluntary; which one might plausibly interpret as meaning that the desires are causal origins in oneself that cause voluntary action.[43] Finally, in the *Rhetoric*, Aristotle lists different

[39] For criticisms of this approach see e.g. Hornsby 2004, Mayr 2011.

[40] MA 7 701a33-36, tr. Nussbaum 1978.

[41] DA III.10 433a17-20, tr. Hicks 1907, modified.

[42] Cf. EE V.2/EN VI.2.

[43] Cf. EN III.1 1111a22-b3.

causes of actions, noting that those actions that are caused by a person themselves are because of (*dia*) desire.[44]

It thus seems that psychological phenomena such as perception, reason, choice, and desire are central concepts in explaining both human behaviour such as actions and productions, but also animal locomotion. Indeed, Aristotle seems to think these *psychological* features explain certain *physical* phenomena like bodily movements. Further, not only do the psychological features *explain* these physical phenomena, but they also seem to *cause* (or originate) them. Finally, these psychological features seem to tell us something about the persons who are acting, and whether they are responsible for their deeds. So understood, Aristotle seems to be advocating some form of a causal theory of action, similar to e.g. Davidson and Smith.

It is clear that Davidson understands Aristotle's views on action as similar to his own. Having made his initial claim that Aristotle is the inventor of philosophy of action, Davidson goes on claim that:

> what is surprising is not Aristotle's interest [in action], or ours, but rather the relative neglect of the subject during the intervening millennia. The reason may be that the connection of action with ethics has been so strong as to overshadow the interest in action for its own sake. But whatever the reason, the consequence is that the subject has progressed, or changed, relatively little since Aristotle. (Davidson 2005, 277-278)

There are two important claims being made here. The first has to do with how Davidson conceives of the subject or topic, and the second is about its history. Davidson seems to distinguish here between "the concept of action" understood as a topic of philosophy on its own – "*for its own sake*" as Davidson has it – which is distinct from ethical considerations about actions. Although he acknowledges that Aristotle was interested with the connection of action to ethics, he credits Aristotle with treating the topics "independently", unlike Plato, whom Davidson claims "was almost exclusively focused on the normative claims on behavior."[45] There are

[44] Cf. *Rhetoric* I.10 1368b28-1369a7.

[45] Davidson 2005, 278.

then two kinds of topics: the nature of actions (considered on its own), and action in relation to ethics. That "action for its own sake" is a question about the nature of actions is clear from Davidson's subsequent claim that most philosophers during the roughly 2300 years between Aristotle's time and the 20th century did not ask "what is the nature of action? but what ought we to do?"[46] And he suggests that framing the question in this way is the reason why the topic has progressed relatively little since Aristotle.

It is a substantial assumption that philosophy of action should be focused on the nature of actions, where this is understood as a distinct question from normative questions of agency.[47] The other assumption is that the topic hasn't progressed much before Anscombe's (and his own) work on the topic. What should be clear at this point is that Davidson's approach to philosophy of action takes the nature of actions as the central question of our topic, that this question is independent from normative questions about agency, and that Aristotle invented *this* way of thinking about the subject. I now turn to the details of Davidson's understanding of Aristotle, which should show that Davidson does not simply mean that Aristotle was interested in the general question about the nature of actions, but also thought Aristotle's approach to the question about the nature of actions is similar to contemporary, causal, approaches.

Assuming for the moment that the subject of philosophy of action is the "nature of action"; what is the way in which Aristotle (and "we") approach this subject? According to Davidson, Aristotle's analysis of action has the following features:

> Aristotle distinguished voluntary actions mainly in terms of the cause: the cause of voluntary actions is internal and mental, whereas involuntary actions are caused by external forces. In the *Categories* he gives as examples of actions cutting and burning; his examples of involuntary actions (also called affects, sufferings, and passions) are being cut and being burned. The cause of

[46] Davidson 2005, 280.

[47] One recent critic of this kind of assumption is Jennifer Hornsby, who raises concerns whether or not this approach contains the resources needed to genuinely accommodate ethical beings, such as human agents and their doings, cf. Hornsby 2004, 2.

> voluntary actions is the conjunction of appetite and thought (*De Anima* 433a). Appetite, which has as its object something valued or desired, initiates the causal chain; thought then determines the means by which the desired end can be achieved. At this point, action ensues. Aristotle stresses that thought alone would never result in action. (Davidson 2005, 278)

Davidson makes several controversial claims here.[48] One is that cutting and burning are (voluntary) *actions*, whereas the *affections* (or passions) like being cut and being burned *are involuntary actions*. Moreover, these are distinguished from each other in terms of their causes: the causes of voluntary actions are "internal and mental", while the causes of affections are external. A further assumption is that the internal and mental causes are the conjunction of desire and thought, where this is understood to mean that desire is the conative starting point of an action, and the thought involved is a means-ends kind of practical thought which helps the agent find the right means to satisfy their desires. Indeed, thought itself is inert; a feature of Aristotle's views Davidson claims is "essentially" like Hume's.[49]

To discuss all of the claims is a task that lies outside my present aim. No doubt Aristotle can been read in the way Davidson suggests.[50] What should now be clear is that Davidson not only thinks that Aristotle is interested in the nature of actions, but also, that the way Aristotle sought to explain this nature is by employing much of the same conceptual apparatus contemporary philosophers make use of. According to Davidson, Aristotle is the father of philosophy of action because he, like Davidson (and others), thinks that explaining the nature of action should be our primary focus, and that this question is best answered by appealing to causa-

[48] An initial assumption that Davidson makes which I cannot deal with here, is whether Aristotle's conception of voluntary (*hekousion*) action corresponds to our notion of intentional action. Settling this question requires discussing if "intentional" is a fitting translation for *hekousion*, how Aristotle uses the concept, and what "intentional" means. For some discussions of these issues see e.g. Preus 1981, Charles 1984, Coope 2010, Price 2016.

[49] Davidson 2005, 281, cf. 279.

[50] I discuss some such interpretations below.

tion, and the mental concepts such as beliefs and desires, which initiate "the causal chain" and which we can use to attribute agency to a putative agent. To evaluate these claims I will focus on the aspects of Davidson's interpretation of Aristotle that he takes to be in alignment with his own philosophy of action. These can be made clear by considering Davidson's claim that the causes of voluntary actions are "internal and mental". What does Davidson mean by "mental"? Further, how do these *cause* action or bodily movements? These are important claims, and how one understands them will affect how one thinks about action.

Perhaps the most striking feature of Davidson's interpretation is that he understands Aristotle as an early proponent of Davidson's own anomalous monism. In his brief discussion on Aristotle's views on the "mind-body problem", he notes that Aristotle "concludes that the mental and the physical are in effect two aspects of the same phenomena", calling this an "enlightened view" since according to it "no problem can arise concerning the causal relations between thought and the physical world".[51] Presumably there is no problem between causal relations because one and the same event can be described in both mental and physical terms, and because relations hold between events no matter how they are described (according to Davidson). That Davidson sees Aristotle's position as similar to anomalous monism is clear from his claim that:

> [...] Aristotle insisted that mental states are embodied, and he claimed that the mental and the physical are just two ways of describing the same phenomena. [...] I applaud Aristotle and Spinoza; I think their ontological monism accompanied by an uneliminable dualism of conceptual apparatus is exactly right. (Davidson 2005, 290)

However, one difficulty here is that Davidson does not elaborate on how he understands Aristotle's ontological monism. What is the basis for this monism? What entities does it contain? Recall that on Davidson's own approach the entities here are events (and objects), which can be described in both

[51] Cf. Davidson 2005, 280.

mental and physical terms, and which function as the causal relata.

In making this claim Davidson appeals to Aristotle's arguments in *De Anima* I.1 where Aristotle discusses the correct way to define affections of the soul such as anger and desire. Aristotle suggests that some might try to define anger as a "boiling of the blood", while others might try and define it as a "desire for revenge". Aristotle concludes that as affections of the soul are related to the body, one should seek to incorporate both (in some way), since such affections are "enmattered" accounts (they are *logoi enhuloi*).[52] The details of this approach are a matter of controversy. Davidson only notes that "Thus definitions of these affections [anger] should define movements of the body (or a part or faculty of the body). The physicist will define anger as a boiling of the blood; the dialectician will define it as an appetite for returning pain for pain."[53] There is then something (anger) which can be defined in physical terms (boiling of the blood) and mental terms (a kind of desire or appetite). One problem is that on Davidson's favoured account "anger" is presumably a mental description of a state which entails an event, and the event can be referred to in both physical terms – "boiling of blood" – and mental terms such as "anger" (or a particular kind of desire). But Aristotle never formulates a theory of events, so it is not clear if Aristotle would agree with Davidson that anger is an event with different descriptions. Perhaps Davidson took Aristotle's activities (*energeiai*) and changes (*kinêseis*) to be events of some kind.

Indeed, Davidson claims that the lack of clarity regarding events is one aspect where contemporary philosophy of action has moved "beyond" Aristotle's. He credits Anscombe with raising the question of how different actions are related to one another, suggesting that some actions are identical:

> This claim immediately raises a number of questions, the first of which is, what are the entities that are identical or different? Actions, we answer; but then, what sort of entities are actions? They would seem to be events. But modern logic had assigned no role to events as serious members of the ontology of the

[52] Cf. *De Anima* I.1 403a25.

[53] *Ibid.*, 280.

> world, nor is it clear how events, particularly actions, are named or referred to in sentences like "Arthur wrote a check." This is an issue of a kind Aristotle was not in a position to discuss with the relative clarity with which it can be raised in the context of today's logic and semantics. (Davidson 2005, 284)

It is a major assumption that we need to introduce events in order to be able individuate actions. For now, suffice it to say that Davidson's interpretation of Aristotle seems to commit Aristotle to accepting a theory of events of some kind, which permit both mental and physical descriptions to apply to them – this is the irreducible, conceptual, dualism – despite Aristotle's silence on this topic.

Davidson's silence in turn further complicates our evaluation of his understanding of Aristotle. As we saw above, not only does Davidson allow events to be described in different ways, but he also takes events to function as causes. However, Davidson does not comment on Aristotle's views on causation, and we are left in the dark on this central point. Nor does he explain how Aristotle's views on causation relate to explanation. Davidson claims that desire "initiates the causal chain" but offers us no insight into what this chain involves. However, since Davidson claims that Aristotle invented philosophy of action "as we now think of it", it would not be unreasonable to assume that Davidson also thought Aristotle's causal model is committed to the same assumptions he is: that the causal relata are spatiotemporally distinct particulars of some kind. However, there are reasons to doubt both the claim that Aristotle's arguments in *De Anima* commit him to a view like Davidson's anomalous monism, and the suggestion that Aristotle's views on causation are similar to Davidson's; I return to these points below.

Even if it is true – as I think it is – to say that Aristotle is interested in the nature of actions there are two things one should consider: what kinds of questions does he raise about the nature of action, and how are these questions answered? I argued above that Davidson takes the question about what the nature of action is to be a basic question in his philosophy of action, as he attempts to explain agency (partially) by determining which events are actions. This is not Aristotle's approach. Indeed, it is not clear whether Aristotle thought that actions are in some sense more basic than agency.[54] One thing we must be clear about here is that "agency" for Aristotle is not restricted to living (rational) substances but is found throughout his philosophy of nature. Substances like the elements and plants are both counted as agents of some kind, with unique causal powers of their own. Moreover, Aristotle's philosophy of nature takes change as a basic principle, and he defines change as an activity (or actuality) of a certain kind of potential.[55] The potential here is something a substance possesses. This suggests that both agents and activities must be given equal consideration for a correct understanding of the ontological basis of Aristotle's philosophy of action. Furthermore, Aristotle's philosophy of nature takes teleology seriously. Although there are disagreements over nearly all aspects of his views on teleology, in general, commentators agree that at least some (if not all) substances have good ends (given their natures), and that they are somehow directed at their ends. "Goodness" is thus a causal or explanatory feature of Aristotle's natural philosophy, which puts pressure on the suggestion that questions about the nature of actions and normative questions are distinct on Aristotle approach.[56]

Indeed, contemporary philosophy of action generally sets itself the task to elucidate what action is, and typically the way one goes about answering this question is the approach Davidson helped inspire. We begin with a neutral category of occurrences, events, and ask what features must be involved for the event in question to count as an action. And the answer to this question that is standardly given within the event causal framework is that the event has to be a bodily move-

[54] Davidson need not think actions are ontologically more basic than agents, but on his approach actions have a kind of epistemic priority, since determining what events are actions also determines if someone is an agent.

[55] See *Physics* III.1-3. The details of this account are a matter of much debate. See e.g. Coope 2007 and Charles 2015 for different views.

[56] Note that although Davidson thinks that there is an irreducible normative feature in the mental vocabulary we use to explain action, normativity is introduced as part of our practice as interpreters and explainers. For Aristotle, goodness is an inherent feature of the world, not something that enters the picture through our explanatory practices.

ment which is caused in the right kind of way by certain psychological phenomena, viz. beliefs and desires.[57] But "what is action?" is not a question Aristotle raises. Even if Aristotle is interested in what distinguishes (human) rational agency from purposive animal motion, and what distinguishes this from elemental motion, the way he approaches these questions is quite different from many contemporary approaches.[58] As I see it, one way this difference is apparent is that Aristotle begins by positing a number of different kinds of actions, changes and activities, and compares them in different ways to show in what way they are different, and in what way they are similar.

Even if few would agree with Davidson's interpretation of Aristotle's views on action, some have nonetheless adopted Davidson's way of doing philosophy of action when discussing Aristotle's views. For example, Alfred Mele suggests that philosophers working on causal theories of action should consider Aristotle because he "presents us with an ambitious and far reaching theory of action *of the sort toward which philosophers are now tending*" – a theory which according to Mele is based in metaphysics and "provides substance" to Aristotle's ethics and moral psychology.[59] Mele develops a causal reading of Aristotle's practical reason similar to Alvin Goldman's. In fact Mele simply refers to Goldman for the details of this theory.[60] Goldman's views, in turn, build on the idea that practical reason involves the use of belief and desire, where belief and desire are characterized as mental events, such as believing and desiring something.[61] While Goldman is critical of Davidson's views on events (he thinks we need to give a more fine-grained account of events), he is not critical of the general events-based, causal, approach to action. Mele thus seems to adopt the Davidsonian approach to philosophy

[57] For a critical discussion of this approach see Ford 2011.

[58] For a helpful discussion on contemporary approaches see e.g. Lavin 2013. Note that although philosophers who are critical of standard, causal, approaches to action, they nonetheless tend to take "what is intentional action?" as *the* question to answer.

[59] Mele 1981, 281; cf. Mele 1984.

[60] Cf Mele 1981, 292n11.

[61] Cf. Goldman 1970, Chapter IV.

of action, without asking whether it is the right kind of framework for assessing Aristotle's views.[62]

Let me raise one general issue with Davidson's and Mele's claims that Aristotle's views on action resemble ours. Aristotle's discussions on action and agency are spread throughout the corpus, and it is not clear if these different discussions amount to a unified view. More problematically, it is not clear if the principles of natural science used to explain e.g. animal locomotion have direct bearing on human action as discussed in the *Ethics* or if these discussions should be treated as distinct topics.[63] Davidson and Mele both assume that Aristotle's metaphysics and natural philosophy provide substance to his ethics; but this is a major exegetical assumption to make (tempting as it may be). However, for my present purposes I will assume that Aristotle arguments made throughout the corpus do rest on each other (and so can be taken to contain a unified theory of action and agency). But even if this is granted Aristotle's philosophical framework is quite different from e.g. Davidson's.

A reason to doubt the applicability of Davidson's framework to Aristotle's philosophy emerges by considering how that framework came to be. As I hope to show next, there are (at least) two important developments in the history of phi-

[62] A more careful approach is found in Charles 1984, who develops Aristotle's philosophy of action in light of Aristotle's own metaphysics.

[63] The *Ethics* might be considered as the closest that Aristotle gets to formulating something like a unified theory of human action (*praxis*). However, in the *Ethics* Aristotle considers *praxis* to be a distinctively human activity – something children and brute animals are incapable of (cf. EN VI.2 1139a18-21). *Praxis* is thus something only adult humans with fully developed capacities for rational thought can accomplish; and it isn't clear if Aristotle think this rational activity can be the subject of explanation in terms of causes. That is: it isn't clear that rational activity is explicable in terms of Aristotle's natural philosophy. Davidson takes Aristotle's discussion on animal locomotion as developed in *De Anima* III.10 as forming the basis for Aristotle's theory of voluntary action, however, voluntary action is a topic primarily developed in the *Ethics* (EN III.1) and as I have just argued it isn't clear if one can apply the natural philosophy to elucidate the points developed in the *Ethics*. This is a difficult question, and answering it lies outside the scope of this paper. Thanks to an anonymous referee for asking me to address this issue.

losophy which affect how philosophers today think about psychology, causation, and explanation. These developments are each, in their own ways, rejections of different Aristotelian positions. Understanding how they are rejections of Aristotelian views should raise doubts that Aristotle's approach to action is as Davidson says it is.

§4. Developments in the History of Philosophy

Two developments following the Early Modern period come to affect our understanding regarding questions in philosophy of mind, causation, and our views on explanation. The first is a form of dualism Descartes' introduced, the other a conception of causation that we find in Hume. Both impact subsequent philosophical thought. The Post-Cartesian conception raises the question of the place of human actions and emotions in nature, and the post-Humean conception in turn affects what kind of causes are acceptable in the explanation of phenomena. These views are further affected by the growing belief that the best explanations we can give are ones in terms of laws of nature – a third significant development.[64]

Briefly put the post-Cartesian conception of mind-body dualism is the idea that mind and body, or mental and physical phenomena, are distinct substances, and given that they are distinct, explanations of physical phenomena are distinct from explanations of mental phenomena. According to Descartes, both mind and body are substances in their own right (or as he has it: "thinking" and "corporeal" substance), capable of separate existence (or at least independently knowable), and both have (at least) one feature or property that belongs to, and only to, the substance in question. For the mind this is thinking, for the body, extension.[65] Since both kinds of substance have an essential attribute through which they are known (and on which their other properties depend on), and since the mind is not essentially extended (or vice versa), it must be the case that mind and body are distinct. Mental phenomena can thus be known without any reference

[64] Cf. Stoutland who notes that the "Davidsonian picture has its roots in the Cartesian revolution" (2011a, 19-20).

[65] Cf. *Principles of Philosophy* I.51-53, 60 (in *The Philosophical Writings of Descartes* ["CSM"] 1:210, 213).

to matter, and corporeal substances can be known without reference to thinking. So understood, physical sciences and psychology are distinct, and one cannot be reduced to the other.

More importantly for our purposes, Descartes also redefines what counts as thinking, and conceives of matter as pure extension, which sets his views on matter apart from his Scholastic predecessors. In the 2nd Meditation we famously learn that the Meditator is only a "thinking thing", but we also learn that this thinking includes a whole range of activities, including doubting, understanding, affirming, denying, willing, imagining and perceiving.[66] By including volitions and perceptions as kinds of thought, Descartes opposes a Scholastic tradition. According to this *Aristotelian* tradition, these activities are split between the rational (human) soul and the animal soul. Further, the soul and body are not two distinct substances, rather soul and body are related as form to matter.[67] Matter (or the body) requires a form in order to function (indeed, in order to be counted as a substance or thing), and the soul, or form, in turn requires suitable matter, with corresponding potentialities to actualize. By locating perceptions (and volitions) as kinds of thought, Descartes alters this Aristotelian picture: all activities of the soul are activities of thought which belong to one substance, the mind (and which cannot be explain physically, since the mental and the physical are distinct).

Descartes also thinks of matter in a different way from his Scholastic predecessors. Lilli Alanen argues that Descartes' notion of matter "defined in terms of extension excluded the entire framework of the traditional philosophy of nature with forms actualizing potencies of material bodies [...]."[68] Descartes' aim is to show that some functions the Scholastics had assigned to the animal soul could be explained mechanistically with the help of the principles of mathematics. Indeed, in *Le Monde*, Descartes defines "nature" *as matter*, whose changes are governed by *laws of nature*,[69] which God has cre-

[66] Cf. *Meditations,* CSM 2:19.

[67] At least for living substances like plants, animals and humans.

[68] L. Alanen 2014, 94.

[69] Cf. CSM I, 92-3.

ated, whereas for Aristotle and the Scholastics, nature is the explanatory principle of changes.[70] Since these laws of nature are explicable in mathematical terms "corporeal nature [...] is the subject-matter of pure mathematics" – as the Meditator concludes at the end of the 5th Mediation.[71] Thus, on Descartes' view, psychology deals with thinking, while corporeal phenomena are to be explained by the principles of mathematics.

These are significant developments, because they amount to a rejection of Aristotle's philosophy of nature and his hylomorphism. Aristotle would not agree to treating nature simply as matter, nor would he agree with treating form (or soul) and matter as distinct substances. Indeed, as Aristotle argues in *Physics* II.1, natural things have both formal and material natures.[72] Moreover, Aristotle would presumably deny that the study of nature is the subject-matter of "pure mathematics", since he argues in *Physics* II.2 that the method of the natural scientist and mathematician differ. The mathematician can treat the objects of explanation in abstraction from any material aspect, whereas one cannot properly explain natural phenomena by abstracting the formal features of a thing from its matter. This is partly because Aristotle's conception of matter is not limited to extension, which is what Descartes takes as the essential property of matter.[73]

Further, these changes raise new issues regarding how we think about human action and behaviour. Indeed, if mental and physical phenomena are distinct, how can one cause or explain the other? This can be called the problem of interaction, and this problem was already put to Descartes by Prin-

[70] Cf. *Physics* II.1 192b20-23.

[71] CSM 2:49. Cf. *Principles of Philosophy* I.64 (CSM I, 247).

[72] See also *Physics* II.8 199a31-33.

[73] Indeed, for Aristotle matter (*hulê*) is the matter *of something*. E.g. wood is the matter of a bed, earthy atoms are the matter of wood, etc. Even if one argued that Aristotle accepted a notion of "prime matter" – which is controversial – this would be matter as pure potency and hence neither extended or corporeal. Aristotle also seems to suggest that premises are the matter of a conclusion – another case where matter is non-extended (see *Physics* II.3 195a15-19; *Metaphysics* IX.7). Thanks to an anonymous referee for raising these points.

cess Elisabeth.[74] Davidson's anomalous monism avoids the problem of interaction, since he rejects substance dualism. Davidson nonetheless accepts that psychology and physical sciences are distinct explanatory projects. However, Aristotle's hylomorphism suggests an alternative approach: to understand natural phenomena we need to invoke the relevant formal and material aspects of it. In "Mental Events" Davidson suggests that "mental and physical predicates are not made for one another", and that there are "no strict psychophysical laws because of the disparate commitments of the mental and physical schemes."[75] But in the context of Aristotle's natural philosophy, formal and material aspects do not belong to distinct schemes. Aristotle might agree that there are no "strict laws" that govern natural phenomena, but his reason for making such a claim is that claims about natural phenomena apply "always or for the most part".[76] Indeed, even those contemporary philosophers who are critical of Davidson's anomalous monism tend to agree that mental and physical concepts are distinct and either that they belong to different explanatory projects, or that mental concepts should be dispensed in favour of physical ones. Aristotle's approach is different, and his natural philosophy makes use of both mental (or formal) and physical (or material) concepts.[77]

To see what significance this has for how we think about action consider the following. For Davidson actions are explained by in terms of reasons, and thus by rational principles. However, given his anomalous monism, our understanding of other's actions and reasons are, in a way, indirect. We do not observe people's actions in the world, but infer them based on our understanding of ourselves. For Aristotle we can observe the goodness of other's actions directly, at least in principle, since goodness is a causal factor in Aristotle's natural philosophy.[78] For Aristotle causes explain

[74] Cf. Letter to Descartes, 6 May 1643, in Shapiro 2007.

[75] Davidson, 2001, 218 & 222.

[76] Aristotle can say this because he thinks that the necessity that holds for nature is not like mathematical or strict necessity. See *Physics* II.2 and II.9

[77] For an in-depth discussion on the distinctive feature of Aristotle's philosophy of mind, see Charles 2008.

[78] The details of Aristotle's theory of perception lies outside my present aim to consider.

phenomena, and in this respect Aristotle may be closer to more straightforward causal theories.[79]

Another development of importance for contemporary philosophy of action is the "modern debate" over causation, culminating in the views of David Hume. Prior to the Modern period, the Scholastics where happy to accept the Aristotelian conception of four causes (with some modifications). This view is drastically altered, following the early modern debate on causation.[80] The significant changes are the following: (i) only the efficient cause is retained as a proper cause;[81] (ii) genuine explanations are in terms of those proper causes (that is: efficient ones); (iii) the cause and the effect are distinct things.[82]

These ideas can be found at work in Hume, and his claims in the *Treatise* are often taken to mark the end of the attempt to reveal the underlying nature of causation, and as a rejection of agency and final causes in favour of "the austere view" that causation is efficient causation, and where efficacy is reduced to constant conjunction with the result that "causation is a matter of brute regularity."[83] Hume's claim that efficacy is a matter of a "constant conjunction of two objects" is certainly influential on later philosophers. While Hume does not seem to have a set view on what he takes the objects (or events) in the causal relation to be, it is clear that they are temporally distinct.[84] More importantly, Hume also denied that there is a conceptual connection between a cause and its effect. According to Della Rocca "causes do not make their effect intelligible. Of course, the cause *together with* certain independent facts, such as regularity or constant conjunction, may, for Hume, explain the effect. But [...] the cause – taken

[79] Cf. *Physics* II.3 194b16-23.

[80] Here too the seeds of change were sowed by Descartes; cf. Della Rocca 2008.

[81] However, see Tuozzo 2014 (23-24) for a discussion on some differences between Aristotle's conception of efficient causes and the Modern one.

[82] For a helpful overview of how the Scholastics' Aristotelian notion of causes and causation gradually came to be discarded in favour of a modern notion, see Clatterbaugh 1999.

[83] Cf. David Hume, *A Treatise of Human Nature*, 1.3.14.32 in Norton & Norton (eds.), 2007; for a discussion, see Kail 2014, 232.

[84] Cf. *Treatise of Human Nature*, 1.3.2.7.

on its own – does not explain the effect."[85] Although Davidson's views on causation differ in certain respects from Hume's, his view is nonetheless *Humean*. Like Hume, Davidson thinks that cause and effect are temporally distinct entities. Further, Davidson also holds there is a distinction to be made between the causal relation and explanation.

But what about Aristotle's view on causation? First off, it is a matter of debate whether Aristotle's four *aitia* should be understood as four different kinds (or modes) of "causes", or four different kinds of "explanation" – although what the differences between these two is supposed to be, is not always clear.[86] Understood as "explanation" this would mean Aristotle's claim that desire is the *archê* and *aition* of locomotion amounts to the claim that desire explain locomotion by being its efficient cause or point of origin, but this is consistent with the claim that the action or motion is explicable in other ways too, e.g. in terms of the goodness the agent sees in the action or movement; and it isn't clear whether either of these is more basic than the other.[87] So understood, explaining animal behaviour is open to different kinds of explanation, and it isn't clear that Aristotle would favour taking explanations in terms of efficient cause as the basis for his philosophy of action or theory of animal behaviour. However, for the purposes of this paper, I will assume that *aitia* should be translated as "causes" – since this translation is more germane to Davidson's reading. However, as I will argue next, Aristotle's view of causation is still different from Davidson's, who is, as I've argued, a Humean.[88]

For Aristotle, the items in the causal relata are not events, but rather substances. Causation is a relation that hold be-

[85] Della Rocca 2008, 236.

[86] Cf. Stein 2011.

[87] Given that what initiates the chain of movers that ends in or constitutes animal locomotion is the good end that is pursued, this suggest that the final *aition* is prior to or identical with the efficient one – in either case the goal cannot be dispensed with if one aims to understand the animal behaviour in question (according to Aristotle). Crucially these are not different explanatory schema that belong to different sciences – both are required for a proper understanding of Aristotle's philosophy of nature.

[88] Many thanks to an anonymous referee for forcing me to develop this point.

tween an agent and a patient, something that undergoes the change or action the agent causes. While Aristotle can agree that the substances in the relata are distinct, he nonetheless argues that the agent's action or activity, and the change or affection the patient undergoes, are numerically the same entity, even if different in account (*logos*).[89] But if the agent's action and the change the patient undergoes are one and the same, then it is not clear in what sense the agent's action can be used to explain the change the patient undergoes, since they are the same. And if they are the same entity then we cannot give a reductive explanation of the change the patient undergoes. But it is not clear if we should even expect Aristotle to give a reductive account of change, given that changes, agents, and patients are basic principles of his natural philosophy. Davidson is right to say that Aristotle is interested in the nature of actions, but actions for Aristotle turn out to be exercises of agency. Moreover, exercises of agency are or involve the exercise of a patient's potential to be affected. Both agency and patiency are at the heart of Aristotle's approach; his philosophy of action is equally a philosophy of passion. The affections ("being burned") and, generally, the effects actions ("burning") have are not distinct ontological categories and thus not distinct entities, *pace* Davidson's claim. As I see it, there is a conceptual connection between the agent's action and the change or affection the patient undergoes: one cannot thus have one without the other. Aristotle can accept that describing an occurrence as an action or activity entails there being a corresponding affection, but this does not amount to Davidson's claim that an action entails there being two events.

While the differences highlighted here do not tell us what Aristotle's approach to philosophy of action is, they do cast doubt on Davidson's claim that Aristotle approached philosophy of action in the same way we do. Even if Aristotle's position on action could be developed along the lines Davidson and others have suggested it would be an anachronistic reading of Aristotle, because Davidson's favoured approach

[89] Cf. *Physics* III.3. Commentators disagree over how the details of Aristotle's claims should be understood. For a discussion see e.g. Coope 2004, Ford 2014, Charles 2015.

is itself a product of different philosophical developments. This does not mean that Aristotle's approach to action does not rely on discussions of psychological features of agents, on the nature of change, action and affection, on questions about the nature of explanation, etc. However, whatever affinity there is between Aristotle and us on these discussions, we should be wary of thinking that Aristotle applies them in the same way we do, or to the same questions we are interested in.

§5. Concluding Comments

In §1 I pointed out that philosophical questions can be understood in different ways – as, say, metaphysical or conceptual ones, and the differences between these is not always sharp. And as I pointed out above, Aristotle's discussions on human action, animal motion, and agency are not confined to a single discussion, but are spread over several different works, all of which focus on different topics: natural philosophy, ethics, psychology, etc. It is thus unclear how, or even if, these different discussions can be brought together to form a single coherent view of action or agency. To complicate matters further, it is an ongoing scholarly debate regarding what kinds of questions Aristotle raises within a single work – whether or not he is, e.g. raising a conceptual question, and what metaphysical implications such questions may have. Saying that Aristotle approached philosophy of action in the way we do simplifies many difficult exegetical questions regarding the kinds of philosophical questions he was interested in, what he took as acceptable answers to those questions, and how the works within his corpus relate to each other.

It should hopefully be clear that Aristotle's claims that desire is the cause of action and locomotion are made within a very different philosophical framework than the Davidsonian framework which many (but not all) philosophers working on action employ. Aristotle may have inspired contemporary work in this field, but he did not share the same assumptions, as his work is grounded in a different ontology and a different conception of causation. His approach remains an interesting alternative, and a continuing source of inspiration to those who seek an alternative to the standard story of action.[90]

It may seem like an obvious point that one should not appeal uncritically to past philosophers as representatives of contemporary views or ideas. Indeed, it may be so obvious to some that the assumption does not warrant a serious objection. This is a pity, as it helps implement the causal theory of action as the orthodox theory, not only in contemporary philosophy, but in history as well. Jennifer Hornsby has raised the concern that if a certain kind of causal account of action is taken as the basis for how questions about ethics and moral psychology are raised then "a shape is imposed on those questions that they should never have been allowed to take on. Meanwhile the orthodoxy of philosophy of mind is silently reinforced."[91] If Hornsby's task is to alert us that contemporary ways of doing philosophy of mind and action might be detrimental to our ways of doing ethics, then my own attempt here has been similar, but with respect to the history of philosophy. I hope to have given some reasons to doubt Davidson's claim that philosophy of action has progressed relatively little since Aristotle.

I believe our history contains quite interesting approaches to philosophy of action, including its nature, and not just how it connects with ethics. But in order to find these approaches, our interpretations of historical views need to be clear on what our current assumptions about this topic rest on. Being alert to these assumptions should help us see in what way past philosophers approached, *or did not approach,* certain questions or problems. This may not only reveal to us novel approaches that may be helpful for finding new ways of tackling or avoiding existing problems, but it is also a more careful approach to the history of philosophy. The history of philosophy of action and agency remains to be written – who knows what surprises it may contain?

University of Oxford

[90] Indeed, many contemporary philosophers have recently turned to Aristotle to find new ways of approaching different issues in philosophy of action. It is a further question whether these approaches adequately represent Aristotle's views on action.

[91] Hornsby 2004, 3.

References

Alanen, L. (2014), "The second meditation and the nature of the human mind", in D. Cunning (Ed.), *The Cambridge companion to Descartes' meditations*. Cambridge : Cambridge University Press, pp. 88–106.

Casati, R. and Varzi, A. (2015), "Events", in E.N. Zalta, (*ed.*). *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition). URL = https://plato.stanford.edu/archives/win2015/entries/events/ (accessed May 2018).

Charles, D. (1984), *Aristotle's Philosophy of Action*. Duckworth: London.

Charles, D. (2008), "Aristotle's Psychological Theory", in *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 24, pp. 1–49.

Charles, D. (2015), "Aristotle's Processes", in M, Leunissen (ed.) *Aristotle's Physics : A Critical Guide*. Cambridge: Cambridge University Press, pp. 186–205.

Clatterbaugh K.C. (1999), *The causation debate in modern philosophy, 1637-1739*. New York: Routledge.

Coope, U. (2004), "Aristotle's account of agency in *Physics* III 3", in *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 20, pp. 201–227.

Coope, U. (2007), "Aristotle on action", in *Proceedings of the Aristotelian Society Supplementary Volume*, 81, pp. 109–138.

Coope, U. (2010), "Aristotle", in T. O'Connor & C. Sandis (*eds.*), *A Companion to Philosophy of Action*. Malden, MA: Wiley-Blackwell, pp. 439–446.

Cottingham, J., Stoothoff, R., & Murdoch D. (*eds.*). (1963), *The philosophical writings of Descartes*. Cambridge: Cambridge University Press.

Davidson, D. (1963), "Actions, Reasons, and Causes", in *The Journal of Philosophy*, 60, pp. 685–700.

Davidson, D. (2001), *Essays on actions and events*. Oxford: Oxford University Press. (2nd ed.).

Davidson, D. (2004), "Problems in the Explanation of Action", in D. Davidson (ed.) *Problems of Rationality*. Oxford: Oxford University Press, pp. 101–116.

Davidson, D. (2005), "Aristotle's Action", in D. Davidson, (ed.) *Truth, language and history*, Oxford: Oxford University Press, pp. 274–294.

Della Rocca, M. (2008), "Causation Without Intelligibility and Causation Without God in Descartes", in Broughton & Carriero (eds.) *A Companion to Descartes*. Oxford: Blackwell, pp. 235–250.

Ehring, D. (2014), "Contemporary efficient causation : Humean themes", in Schmaltz (2014), pp. 285–310.

Ford, A. (2011), "Action and Generality", in Ford, Hornsby, & Stoutland (2011), pp. 76–104.

Ford, A. (2014), "Action and Passion", in *Philosophical Topics*, 42, pp. 13–42.

Ford., A., Hornsby, J. & Stoutland, F. (eds.) (2011), *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press..

Goldman, A. (1970), *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall (reissued Princeton, NJ: Princeton University Press, 1976).

Hicks, J. (1907), *Aristotle. De Anima*. Cambridge: Cambridge University Press.

Hornsby, J. (2004), "Agency and actions". Hyman & Steward (2004), pp. 1–24.

Hyman, J., & Steward, H. (*eds.*) (2004), *Agency and Action*. Cambridge: Cambridge University Press.

Kail, P. (2014), "Efficient causation in Hume", in Schmaltz (2014) , pp. 231–257.

Lavin, D. (2013), "On the Problem of Action". URL = http://nrs.harvard.edu/urn-3:HUL.InstRepos:9887629 (accessed May 2014). Published in German, in *Deutsche Zeitschrift für Philosophie*.

Mele, A. (1981), "The Practical Syllogism and Deliberation in Aristotle's Causal Theory of Action", in *The New Scholasticism*, 55, pp. 281–316.

Mele, A. (1984), "Aristotle on the Proximate Efficient Cause of Action", in *Canadian Journal of Philosophy Supplementary Volume*, X, pp. 133–155.

Mayr, E. (2011), *Understanding Human Agency*. Oxford: Oxford University Press.

Norton D. F., & Norton M. J. (eds.) (2007), *The Clarendon Edition of the Works of David Hume. A treatise of human nature.* Oxford: Oxford University Press.

Nussbaum, M. (1978), *Aristotle's De Motu Animalum*. Princeton: Princeton University Press.

Price, A. (2016), "Chocie and Action in Aristotle", in *Phronêsis*, 61, pp. 435–462.

Preus, A. (1981), "Intention and Impulse in Aristotle and the Stoics", in *Apeiron*, 15, pp. 48–58.

Schmaltz, T. M. (ed.) (2014), *Efficient causation : A history*. New York : Oxford University Press.

Shapiro, L. (ed.) (2007), *The correspondence between princess Elisabeth of Bohemia and René Descartes*. Chicago; London: University of Chicago Press.

Smith, M. (2004), "The Structure of Orthonomy", in Hyman & Steward 2004, pp. 165–194.

Smith, M. (2012). "Four objections to the standard story (and four replies)", in *Philosophical Issues*, 22, pp. 387–401.

Stein, N. (2011), "Causation and Explanation in Aristotle", in *Philosophy Compass*, 6/10, pp. 699–707.

Stoutland, F. (1989), "Three conceptions of action", in H. Stachowiak (*Ed.), Pragmatik : Handbuch pragmatischen denkens.* Hamburg: Meiner, pp. 61–85.

Stoutland, F. (2011a), “Anscombe's intention in context”, in Ford, Hornsby, & Stoutland 2011, pp. 1–22.

Stoutland, F. (2011b), “Interpreting Davidson on intentional action”, in J. Malpas (*Ed.*), *Dialogues with Davidson : Acting, interpreting, understanding*. Cambridge, Mass.: MIT Press, pp. 297–324.

Tuozzo, T. (2014), “Aristotle and the discovery of efficient causation”, in Schmaltz 2014, pp. 23–47.

On the Homeostasis of Virtue

ANTTI FENIX SNEITZ

Virtues cluster. This was known to Socrates and Aristotle. In fact much in Aristotle hangs on it. Yet, modern revival of virtue ethics tended at first to look past the issue or assume that no such unity was to be found. The issue which in many ways was central to Greeks seemed "to have dropped out of consideration" as Cooper observed (1998, 76). In a similar vein Badhwar claimed that "most commentators on this doctrine have tended to dismiss it" (1996, 306). Was that not merely a curiosity of the Greek way of thinking or some half-thought mystical ideal of sage-like perfection? Why were the ancients so keen on virtue being one? And why should we be?

Different accounts of virtue give rise to different reasons for unity. The picture we find in Aristotle, however, is particularly riddling. For him moral virtues are apparently many, and moreover, there are intellectual virtues, distinct from these. Yet Aristotle and his followers too insisted that virtues mutually entail each other. The solution to this riddle is found in the peculiar dependence between the two sorts of virtues in Aristotle. Intellectual virtues are brought to bloom only by the antecedent development of moral virtues, but genuine virtue is achieved only when the intellectual virtue of *phronesis* (typically translated as "practical reason") has been thus constituted.

Not all virtue theorists vouch for unity. Protagoras, Bernard Williams, Philippa Foot and Von Wright were explicit disunitarians. Protagoras aside, the assumption of disunity was more pronounced in the earlier phase of contemporary discussion. Yet disunity has been more assumed than argued for. Recently, unity, or reciprocity, of virtue has again become a live matter. Number of recent authors have taken up the question (see Annas 2011; Badhwar 1996; Hurtshouse 1999; Sreenivasan 2009). I will offer a take different from these authors in tackling the question of unity with resources provided by Richard Boyd's metaethical theory (Boyd 1988; 2003a, 2003b).

The present paper has two closely connected aims. It gives an account of the unity of virtue applying Boyd's (1988) idea of Homeostatic Property-Clusters (HPCs). On this basis I outline a framework for virtue theory. I will argue that the case is in fact better for virtue than for Boyd's original consequentialism. There are reliable inductive generalizations from virtues supporting inference from one to another, and in taking virtue into account, the clustering of human goods becomes more plausible.

On the Boydian picture clusters of properties are to be united by a mechanism. So if virtue is an HPC kind, what plays the part of a mechanism responsible for the clustering? Finding an answer to this question provides a starting point for sorting out the individual cases of dependence between virtues. However, if a fairly strong naturalist programme (such as Boyd's) is to be assumed, this should not be an *a priori* matter. That is: we should not expect a mere analysis of the concept of virtue to yield an answer.

Aristotle, I believe, was on the right track even if his insistence for overall mutual entailment between virtues is too strong. Unless carefully stated, his approach also runs an obvious risk of circularity. How is anyone to be virtuous if *phronesis*, a precondition of virtue, is only attained by already having mastered the virtues?[1]

In the first two sections I will discuss the notion of virtue and the various forms of unity of virtue in general terms. Third section outlines Boyd's theory. After this, I discuss arguments by Rubin against Boydian naturalism. I will then proceed to treat the unity question in Aristotle and in related traditions in some length. In sections 5–8 I go through how HPC unity of virtue works in more detail. Finally, in section

[1] A puzzle more particular to Aristotle was discussed by Alexander of Aphrodisias: how can virtues form a genus if destruction of one member destroys the whole?

9, I briefly sketch a fusion of the so-called natural goodness account and homeostatic-cluster view.

From this I will be in a position to make explicit a number of details of the homeostatic clustering of virtues. This provides a start toward an attractive, naturalistically acceptable account of virtue, and corroborates particularly well with the biological and psychological foundations of virtue theory. The later part of the paper also advances suggestions based on this discussion concerning Boyd's original metaethical picture.

1. Virtue

A virtue is generally taken to be a trait of character, or an agent's dispositions to act well. The Greek word for virtue is *arete*, and readers are commonly reminded that the connotations of the current counterparts for the word tend to sound more prudish than intended. More could be said on virtue's dispositional nature, or on what exactly is meant by calling virtues dispositions, for example how to distinguish them from skills (see Annas 2011 and Von Wright 1963 for some troubling over). I will follow Hutchinson (1986, 35–36) in not worrying, and equate "disposition" here with the present standard usage.

Virtues are items which can appear within any kind of moral theory. For example, traits or dispositions can tend to maximize utility, or they can be seen as internalized duties. Boyd's consequentialist account does not fail to mention character traits.

But besides the two "modern" re-constructions just mentioned, there is an older tradition of theories based centrally on the notion of virtue. During the past 50 or so years, such accounts have seen a large-scale revival and are now well-known. Most of these, though by no means all, look back to Aristotle in name or in detail. In consequence "neo-Aristotelian virtue ethics" has become something like a rather loose extended family of theories.

Virtue theories tend toward being naturalistic in the metaethical sense. This is by no means a necessity, but both tradition and contemporary interest have strong affinities with a naturalist outlook. Adams is explicitly supernaturalist in his outlook; Aristotle in *Eudemian Ethics*[2] appears more inclined toward something like this, too (see *EE*, 1217a20-26; see also Adams 2006, 49–50). But for most neo-Aristotelians, virtue is a recognizably worldly object, as it were, and unless one sets for it a supernatural standard (like Adams does), it does not require more than psychological objects for its ontology. Virtue does not reside in heaven (though it might only be fully expressible in heaven), nor does it require a strange and alien faculty of intuition to know matters pertaining to it.

One key element in Aristotelian version of virtue ethics is the notion of goodness as functional, and in Foot's term "species dependent" (Foot 1994, 163). Aristotle's notion of good is not univocal (*NE*, 1096a11–1097a23; *EE*, 1217b–1218b): "Each thing strives for its own good" (*EE*, 1218a31). Goodness then is a matter of satisfying certain criteria set by what kind of thing the thing evaluated is:

> Because a cloak has work to do, there is such a thing as the goodness or virtue of a cloak, that is, the best state for a cloak to be in. So too with a boat, a house, and other things. The case is the same with the soul, for it too has work to do. (*EE*, 1218b38-1219a5.)

Virtue is a matter of excelling at something, and thus requires that this something has a purpose or an end, a final cause. Artefacts, like cloaks, have purposes as they have been made for them. But do other things have such ends? Especially, do humans have some such purpose for which they could be said to be more or less excellently suited? Some react adversely to such suggestion, while others deem the very idea of this kind of teleology dead, or at any rate out of place in moral theory, claiming that Aristotelian stance sinks with its associated assumption of natural teleology.

Not just any *ergon* present in man is of interest to ethics as such. Aristotle's ethics follows his tripartite psychology. The lower portion of the irrational soul, the vegetative part, has

[2] Works of Aristotle will be cited by Becker numbers, and the name of the work will be indicated by abbreviation. NE=Nicomachean Ethics, EE=Eudemian ethics, Cat.=Categories, Met.=Metaphysics, PA=Parts of Animals, Phys.=Physics. Except for the Eudemian Ethics, all translations are from Barnes 1984 (ed.).

its way of functioning properly, but this is irrelevant to ethics; only that part of the irrational soul which can be under the control of reason matters. Part of the irrational soul is "capable of following reason, in line with reason's ability to command" (*EE*, 1220b6–7). We do not reason and decide that we are hungry, though may decide to eat. And we definitely do not digest by reasoned decision. This desiring "animal" part of the soul can relate to reason in three ways. In some cases it overcomes reason's guidance (resulting in incontinence), in others, the animal part obeys (resulting in continence). But in the genuinely virtuous the parts work in harmony, so that there is no coercion between the parts (*NE*, 1102b14-1103a2).[3]

There are goings-on in ourselves of which some are under our control and others are not. Among these, there is reasoning, which can (if things go right) harmonize with what the part (more or less) under our control does or is inclined to do. Only that part which in principle can come under reason's command is of ethical interest, and that is where ethically significant virtue resides. But nonetheless, both this ethical goodness and the goodness of the parts which lie beyond reason's reach are of the same functional nature.

In contemporary discussion, it is commonly said that virtues are dispositions. This does not stray too far from Aristotle, who said that virtue is a state of soul, but meant by this something very similar. This should come with a warning that in Aristotle what is commonly translated by the word "disposition" is something quite different, namely arrangement of parts (Hutchinson 1986, 9–10; *Met*. 1022b1–1023b22). But the word used to tell what kind of objects virtues are, "state" (*hexis*), means here a specific kind of property, characterized in a way which turns out to be not too far removed from the contemporary meaning of disposition.[4] A *hexis* is distinct from two other kinds of qualities that can also be found in soul, emotions and capacities. Emotions are passing

[3] Aristotle wavers on whether the animal part is properly grouped into rational or irrational (*NE*, 1103a3; *EE*, 1219a28 versus 1220a10; see Kenny 1978, 167.) This suggests that the distinction is less rigid than might be assumed.

[4] In *Categories* we find that Aristotle distinguishes four sorts of qualities: states, capacities, affection and shapes (*Cat.* 8b25–10a26).

affections, while capacities are what are needed in order to experience these. A *hexis*, on the other hand mediates affections, putting us "in a good or bad condition with respect to the feelings" (*NE*, 1105b25–26). This, I suppose, could strike one as quite narrow. But the idea is that affections cause behavior, and virtues filter these impulses, so that good action comes about.

Do virtues require each other or do they rather support each other? Or is it this way here, and that way there? Let us say that an account of their unity is strong if they are dependent for their existence on the others, and weak if the presence of some makes the appearance of others more likely. More detailed dissection is given presently, and the Boydian apparatus introduced in section 4 will be used to sort out the situation further.

2. Unities and disunities of virtues

The idea about the connectedness of virtues comes in many guises.[5] First a convention to navigate the plenitude: I will call any form of connectedness between virtues "unity" and will distinguish between identity and reciprocity as two kinds of ways of treating this unity.[6]

Different kinds of unity theses can be classified according to how strong unity they claim. In its strongest form the thesis is that there is literally only one virtue. Let us dub this identity thesis. According to it, seemingly different virtues are in reality identical with some single underlying virtue. This position was held by Socrates, who thought that virtue was actually knowledge. The strength of the identity claim comes in degrees, however. Strongest identity thesis is at difficulty in explaining why there are apparently different virtues in the first place, why they come with such different descriptive contents, or why some plausible attributions of one virtue do not evidently entail attributions of all the others. One could suggest that different virtues are all different

[5] Hursthouse (1999, 118) mentions Timothy Chappell having counted 30 or so versions. We'll do with less.

[6] Some authors, like Sreenivasan, designate same distinction by tags "Aristotelian" and "Socratic" (2009, 197–198).

aspects of one condition.[7] But let us take on Aristotle, who is somewhat more moderate.

According to Aristotle, too, there is unity, but instead of identity, virtues are related by reciprocity: an agent cannot have just some of the virtues. They come as a package, or mutually imply each other. Vlastos (1972) called this "biconditionality". I will instead say that biconditionality of virtue is an extreme form of the more general connection type for which I reserve the name "reciprocity of virtue". Weaker forms of reciprocity are also possible, and so are cases where some connections are biconditional, while some others are of a weaker kind.[8] But even in the strong biconditional case, the Socratic collapse of virtues into one is not implied: each virtue is a different property in agent's psychological make up, even if it necessitates the presence of others. It is thus not an identity thesis of virtue(s).

Reciprocity thus grants that different virtues are distinct objects having, say, their own identity conditions, but at the same time claims that they are connected, and indeed in the paradigmatic case, that having of each entails having all the others. Strongest form of this kind of thesis claims necessary entailment between virtues. This is the biconditional reciprocity of virtues. Weaker forms have weaker connections. Reciprocity divides further into direct and indirect accounts. Direct reciprocity runs between any two virtues, but indirect reciprocity of any two virtues comes about due to some underlying third factor. Aristotle's position on unity of moral virtues is of the latter kind, because in his theory, individual moral virtues come as a package due to their dependence on *phronesis*. But beside this distinction, there are also weaker levels of reciprocity, involving, as it were, a statistical entail-

[7] The Socratic account leaves it open whether seemingly separate virtues are aspects or parts of some one underlying thing, or not really separable at all save for example per something like a Fregean *sinn.* (See also Cooper 1998; Vlastos 1972).

[8] And obviously one could suppose that sometimes even a limited identity of some seemingly independent virtues could occur side by side with the reciprocal connections, for example assume that some set of *prima facie* independent virtues v_1, v_2, … are at the bottom really one just one condition, V_i, and that V_i is biconditionally connected to another condition (or cluster thereof), the identity of which is not determined by the V_i. Etc.

ment rather than strict biconditionality. In such case having a virtue makes the appearance of others more likely.

As observed two reciprocities come in two sorts, depending on whether the reciprocity is direct between virtues or whether it runs through a centralized mechanism exhibited by a particular virtue, and these both are further divided according to the strength of this connection. This will be explained in more detail below, as it will be central in my Boydian reconstruction of virtue. (The mechanism could in principle be something else and not itself a virtue, but I will omit that case from discussion).

But aside from all this hairsplitting, to many it would seem natural to assume that virtues can be possessed independently from each other. Are not some of us lecherous while remaining courageous? Or if it is granted that one cannot have a virtue if one also has an actual vice (such as lecherousness), could not the trait of courage then be possible in the mere absence of the virtue corresponding to that virtue? The biconditionalist picture Aristotle gives seems to demand too much in binding the possibility, for example, of being just or courageous with such seemingly unrelated traits such as plausible wittiness (*NE*, 1128a1–1128b12) and not walking too fast (*NE*, 1123b15–20). Surely one can be both just and hasty.

Some philosophers have indeed thought that no unity exists, or that it is at best a very weak statistical matter. It could be said that, in the way I have drawn the distinctions above, weak forms of reciprocal unity fade into disunitarian positions. Connectedness of virtues is a matter of degree, where their biconditionality is the upper limit and absolute disunity the lower.[9] I suggested that studying the various modes of connectedness which lie between these extremes is an important task, in which the homeostatic property cluster conception (introduced in the next section) becomes helpful.

Others, like Philippa Foot (1978, 14–18; 1983, 42–43) have gone even further, suggesting that some virtues are actually contradictory, and so not only is it not necessary that virtues come as bundle, it is impossible that all of them do. That is, some are actually contradictory. (This sort of thing of course

[9] Where in this continuum the line between weak reciprocity and disunity lies is not unrelated to the extensional vagueness of HPCs; see below.

has been puzzled over in the tradition, because of the apparent incompatibility of certain virtues of Aristotle.) The claim calls for some scrutiny, as not all virtues should be expected to be mutually compatible – virtues of fishes need not to be compatible with virtues of men, and even among humans, some may be bound to specific roles or phases of life. Only virtues of sufficient ground in basic human nature should really count here. And it is less clear that these can be in conflict, at least when properly developed.

I will argue that there is sort of a reciprocity of virtue, and that Aristotle was basically right about the way it comes about. But he erred in treating the reciprocity as the strong, biconditional, kind. Rather, virtues form what Richard Boyd has called a homeostatic property cluster. In such cluster, relatively permanent properties exist in conjunction, held stable by a mechanism uniting them. Already this characterization is remarkably similar to many things Aristotle said about virtue: they are relatively permanent states (*hexeis*), which exist in conjunction (are biconditional, entail each other) and they are united by certain functioning of the rational part of the soul.

3. Boyd's conception of Homeostatic Property Clusters

In his widely cited paper (Boyd 1988) Boyd gave an account of natural kinds based on the notion of homeostatic clustering of properties.[10] Homeostatic clustering is law-bound co-occurrence of properties. As Boyd puts it, in such clustering "[e]ither the presence of some of the properties in *F* tends (under appropriate conditions) to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties in *F*, or both" (1988, 329), where *F* is a family of co-occurring properties.

Thus, such items are more than mere statistical clusters. To be a natural kind, said cluster must have a unity provided by what Boyd calls "homeostatic mechanism", a mechanism binding together the co-occurring properties. Although the notion of a homeostatic property cluster (or HPC) is more

[10] This is further developed in 2003a and 2003b in the context of replying to Adams 1999.

generally applicable, here Boyd had in mind a particular metaethical application: certain notions, such as *Human Good* and *Moral Good* were to be treated as HPC kinds (1988, 329–331).[11] In Boyd's view, an explicit continuation with natural science is intended: "I mean the analogy between moral inquiry and scientific inquiry to be taken very seriously" (Boyd 1988, 330).

Boyd's account of natural kinds as HPCs thus gives them much flexibility, but it is also intended to do a job for which a more rigid notion of natural kind is often summoned, namely to function as the ground for persisting referent in a causal-historical theory reference for kind terms. Such theories, as pioneered by Kripke, Putnam and others need to postulate natural kinds to account for the reference of kind terms in parallel with their treatment of logically proper names.

In Boyd's view ethical terms denote such homeostatic clusters: "I think that ethical goodness is probably defined by what I call a homeostatic property cluster: a family of properties of actions, policies, character traits and the like which are aspects of, or contribute to, human flourishing and which are such that they exhibit a sort of homeostatic causal unity: under suitable conditions their instantiations are (causally) mutually supporting" (2003a, 510). *Moral Goodness* is a natural kind consisting of the "cluster [of *Human Goods*] and the homeostatic mechanisms which unify them" (1988, 329).

Human flourishing, or the cluster of *Human Goods*, is assumed to consist of various psychological, social and physical and medical properties. Furthermore, Boyd is also claiming that this clustering is homeostatic. (Boyd 1988, 329–330). In principle, there are two ways for such homeostasis to come about, namely i) mutual support between various individual *Human Goods* or aspects of human flourishing or ii) an underlying mechanism or set thereof.[12] Boyd's account is consequentialist in the sense that the *Moral Goodness* of actions,

[11] Names of putative HPC kinds will be written with capitals and italicized.

[12] This two-part characterization is somewhat redundant, as there will in any case be a mechanism or set thereof to account for the "mutual support". The distinction appears to be between what could be termed centralized and distributed mechanisms for clustering.

policies, character traits, institutions and other suitable bearers consists of their bringing about these human goods.

Boyd's account also allows for imperfect homeostasis. Some clusters can lack some of the properties. Some other characteristics Boyd lists include: HPC kinds lack analytic definitions; determining the relative importance of different portions of the cluster is not a conceptual issue (at least not entirely); unresolvable cases of extensional vagueness are possible; and HPCs can change.

Let us call any metaethical position making use of homeostatic clustering plus realist semantics *HPC Moral Realism*. How to fit virtues into the picture? Boyd does not say much about character traits. I will give an account of virtue as a HPC kind. I will also argue that Virtue, understood in HPC terms, fares best as a candidate for a key property in HPC Moral Realism. Boyd's original notion of *Good* as a HPC cluster suffers from the disunity of its purported object. However, much of this disappears when the good is centered around virtue. Furthermore, virtues provide a natural and much cited example of clustering, one which I argue should be seen as homeostatic.

Before embarking further, let us outline how this treatment would work with virtue. First of all, virtue is an object well suited for the causal job expected of a HPC. In other moral terms, particularly with abstracted ("Moorean") notion of good, the causal efficiency of the denoted object seems suspect.[13] Even in the consequentialist picture, causation involved is indirect, coming about through decisions of agents to act in a certain way. But virtue by its very nature is a causal thing, a disposition in an agent, responsible for causing the agent to react in certain ways under certain conditions. Also, there are, as I will argue in more detail below, causal connections between virtues, sufficient to ground parts of the clustering making up the HPC. Virtues form a system in which these connections play various causal roles.

[13] For Moorean good, see Foot 1994, 162; 2001, 2–6. The argument from causal inefficiency of abstract objects here implied (but not developed) is intended to be analogical to Benacerraf's line of thought about mathematical truth.

Virtues are rather something described materially in the context of ethical theory, and Aristotle even gives a well-known warning against expecting more exactness than the subject matter allows (*NE*, 1103b27-1104a9). Furthermore, determining the relative importance of different virtues in the cluster formed by them is not an entirely conceptual issue, but would seem to require at least experience (*NE*, 1141b8–1142a21). It is likewise plausible to say that virtues often exhibit some "unresolvable extensional vagueness", for example making it sometimes impossible to determine whether a threshold of, say, courage has been achieved. Aristotle's original account however clashes with a number of Boydian suggestions, such as the presumed mutability of HPCs.

This conflict can at least partially be resolved, and it leads to the interesting question about how and how much virtue can change. I will not treat this in detail. Surely, Aristotle presupposed that it can't change. But we do not need to follow him here, as we do not share his assumption about the eternity of the species. With the species relative natural goodness account in hand, post-Darwin, it easy to understand how and why virtue, too, would change.

In the next section I will rephrase arguments against Boyd's position by Rubin (2008). For our purpose, Rubin's arguments will work as test cases. While Rubin's criticism is partially misguided, his arguments point to issues where the virtue HPC realism fares better than Boyd's eclectic consequentialism.

4. Difficulties of HPC moral realism

Rubin (2008) argues against Boyd's HPC moral realism. His attack is based for a good part on confusing Boyd's notions of *Moral Good* and *Human Goods*. As the name suggests, *Human Good* is basically a list of welfare conditions for humans. *Moral Goodness*, on the other hand, is the key notion. This is the cluster of the *Human Goods* plus their unifying homeostatic mechanisms (Boyd 1988, 329; see also Adams 1999, 60–61). The later comprises actions, traits and policies which tend to bring about or maintain the *Human Good*. This is less innocent than Rubin's discussion makes it appear. Rubin says he is restricting his attention to what he calls "non-instrumental

Moral Goodness", and by this he evidently means Boyd's *Human Good* (Rubin 2008, 504–506).[14] But that is not *Moral Goodness* at all, but a rather a list of well-fare items on which the homeostatic mechanism of *Moral Goodness* was to operate, namely the *Human Goods*.[15]

The argument is not entirely mistaken, as Boyd does claim that Human Goods, too, cluster (Boyd 1988, 329), but appreciation of Boyd's position should make it clear that the clustering here is less essential for his picture than the clustering of *Moral Goodness*, and the clustering of *Human Goods* as influenced by the mechanisms of *Moral Goodness*. (If the mechanism was cut from the picture, properties making up *Tiger* wouldn't show much clustering either. There is nothing in stripes to make them come with sharp teeth if both the genetic apparatus and the environment of adaptation are bracketed off.)

Nonetheless the arguments are of interest, as it can be pointed out that HPC virtue theory fares particularly well against them. Rubin's arguments boil down to two, isolated goods argument and family of structural failures arguments (count them any way you like). Isolated goods argument is simple: there could be goods isolated from all clusters. Such possibility should show that *Good* does not need to be bound by homeostatic mechanism (Rubin 2008, 509–513). The argument is rather weak, as it does not take much ingenuity see to how Boydian could avoid it – easiest way is deny the plausibility of each such putative separation scenario. At any rate it concerns the reference of the word "good", presumably showing that something is good, while not being part of a homeostatic cluster. But as I am here interested in clustering of virtues as material phenomena, the analogical argument would merely recapitulate the already cited observation (the

[14] Witness for example the rather peculiar statement: "it is proper to say of John's being in love with Mary that it is morally good" (Rubin, 513).

[15] Boyd should probably have avoided suggesting that *Human Goods* are included in the *Moral Goodness* in the first place. In his picture nothing is gained by this conflation. It would be better (and more in line with the actual usage) to say that the *Moral Goodness* is just the homeostatic mechanism bringing about and maintaining the *Human Good*.

disunity thesis) that sometimes virtues appear to exist in isolation from each other.[16]

The more important structural failures arguments show that *Good* does not behave the way one would expect of HPC. First argument is that individual clustered properties are not normally sufficient for predicating the kind term. The kind *Tiger* is not properly attributed of any of the individual tiger-making properties. But in the case of *Good*, predication is often done just on such basis, as each item on the list of *Human Goods* is presumably an instance of *Good*. (ibid., 513–514) Second is that individual instances of kinds should have all or most of the defining properties, whereas Boyd's *Good* fails this. Tigers come as complete or almost complete bundles of tiger-making properties. It does not make sense to say that something with, say, half of these properties is a tiger. *Goodness*, however, does not seem to require such approximation to completeness. (ibid., 514–515). These arguments are obviously conversions of each other. The point in both is that *Good* behaves anomalously as a HPC kind term.

Rubin also puts forth variations of these two arguments to show analogical failures of inductive generalizability on *Good* (517ff). These are especially poignant given Boyd's insistence on the inductive grounds provided by the nature of clustering. The arguments are: i) that from one property belonging to the putative Goodness-cluster one cannot reliably infer the presence of others; and ii) that from knowing that thing is characterized by this cluster, one cannot reliably infer what properties it has. Compare this again with the case of the tiger – from some properties of a tiger you are in position to infer to other via inductive generalizations. And from the knowledge that something is a tiger, you can infer to its characteristic tiger properties. (Rubin 2008, 519–521).

I think this shows that whatever clustering is present in what Boyd calls *Human Goods* does not probably by itself form a natural kind – or even if it does, per Rubin's first set of structural failures, it is not properly denoted by the word "Good" used as a kind term. So far so good. But I also think it

[16] Observe however that the point could be made against the more general natural goodness account discussed in later sections.

is not essential to Boyd's picture that it should form such a kind.

Now a more detailed reconstruction of *Virtue* as HPC kind will be given. I will start by considering how virtues could in principle work in the framework of HPC moral realism. I will then go on to outline a neo-Aristotelian theory which avoids the problems raised by Rubin.

5. Two plus some ways

There are a number of ways to develop an ethics of virtue from the HPC point of view. First, one can seek to integrate virtues into Boyd's metaethical picture, keeping his basic idea about goodness intact. Developed in this fashion, an account of virtue would supplement Boyd's consequentialist notion of *The Good*, building on his remarks on character traits. The other way is to be imperialist: that is, to throw aside the consequentialist approach and instead make virtue the central, or indeed the sole, notion operative in Moral Goodness. Thus from Boyd, one would take only the HPC conception of basic moral kinds, and leave aside his sketch for a consequentialist theory.

The supplementary approach splits into two: on the one hand, virtue would appear alongside other *Human Goods*, and on the other, it would be something of a matter of *Moral Goodness* as well. It should be noted, that this does not automatically entail treating virtue entirely in consequentialist terms (like Harman 1983, Driver 2001). This is because virtue as a *Human Good* can be non-consequentially understood as natural goodness, but this does not preclude its being operated on by the consequentialist mechanism envisioned by Boyd; after all, promoting virtue surely is good (whatever else may also be). But on the level of Boyd's *Moral Goodness*, the measure of virtue would on such an approach surely be consequentialist.

Putting virtue on the list of *Human Goods* may seem like a strange thing to suggest. But *Human Good* is a cluster made basically of what makes humans do well, and its content is at least largely an empirical question. Concerning this I'll put forth an Aristotelian hypothesis: *human being is not really well if he or she does not have virtue.* Even if all the external goods are at play, a man will not flourish unless he also has a good character.

This can be supported in the following fashion: a man who enjoys plentifully the other *Human Goods*, but does not have proper moral character could suffer due to demands of *Moral Good*. Assume that he is not amoral, but is either continent or incontinent in the Aristotelian sense. What morality demands of him does not come naturally because his desires are in conflict with it. But this results in suffering. This would not happen if he had a good character, in which case he would instead choose the right action without any internal conflict.

On the alternative imperialist approach one would make virtue the sole *Human Good*. This position is attractive for reasons of parsimony, and due to its easing considerably the referential burden of Boyd's sketch. As shown by Rubin's arguments discussed earlier, Boyd's notion of *Human Good* appears stretched too wide. (On the other hand, there would appear to be other things humans need to do well, and this did not escape Aristotle's notice.)

Do I need to choose in which way to proceed? For the large part I think the answer is no. A substantial fragment of a HPC account of virtue can be developed while remaining neutral about the rest of Boydian metaethics. Virtue can for example be treated solely in reference to its function as a *Human Good*. There is a further crossroads however, at the level of *Moral Goodness*. Here the Boydian picture makes the assessment of character traits a matter of their consequences. At this point, the discussion moves onto larger waters in virtue theory, where a number of consequentialist accounts have been put forth. Such construals aside, one could accept that virtues always bring benefits, and it would seem that here indirect benefits would be good enough a compromise. Likewise, granting that virtue is also a basic *Human Good*, not unlike health, makes virtue a morally significant category even if virtue would not provide any benefit external to itself. But unlike health, virtue is not merely a good to be had. It is also operative in bringing about *Goodness*.

Why then flirt with imperialism when such high compatibility is readily obtainable in any case? One reason is that the concept of good inherent in the notion of virtue, the functional natural goodness, provides a superior alternative to consequentialist treatment of *Good*, and can in fact incorporate a good bit of what is attractive in the consequentialist HPC account via the quasi-consequential virtues such as benevolence.

Leaving the question of supplementation versus imperialism aside, I'll proceed to see how the HPC conception works with virtue. After this discussion, some further remarks will be made on the above topic.

6. Connections between the virtues

Considering the then much discussed question whether moral virtues are connected, Duns Scotus remarks: "This question encompasses a number of topics: (1) the connection of the moral virtues with each other, both in terms of their genera and in terms of the species of those genera; (2) the connection of each moral virtue with prudence" (*Ordinatio* III, d. 36, q. un., 10). Omitting here his insistence on the so-called theological virtues and their respective roles, this distinction provides a plausible starting point. Moral virtues, or virtues of character are virtues of the irrational part of the soul, while what is here called "prudence" is the intellectual virtue of *phronesis*.

Scotus, it turns out, is right in that the two kinds connectedness are significantly different. Sometimes virtues depend directly on each other, like in the time-honored Socratic example of piety being needed to be just (for without piety one would lapse into injustice toward gods; see Cooper 1998, 86–87). The virtue of courage is another example, as it provides guts to do what needs to be done.[17] These are cases where the proper functioning of a trait requires proper functioning of another. Let us call such dependencies specific. Given Boyd's rather loose criteria for clustering, such cases are all it takes to posit some level of homeostatic clustering among virtues. This kind of connection is direct between two virtues, and typically it would characterize only some instances of their

[17] Martial nature of courage is sometimes unduly stressed (see Geach 1977, 2–4, for discussion of Hare on this). But courage should be seen as more universally applicable notion. Annas (2011) mentions terminally ill people as exemplifying courage; Geach (1977, 4) remarks that it is needed in urban cycling.

operation. Now we should proceed to ask whether there is a general homeostatic mechanism to be found in the background, explaining and uniting the virtues over and above such specific dependencies.

I will argue that there is, and that Aristotle recognized the matter correctly. Not only do various moral virtues depend on each other, but they all depend on something else, namely the intellectual virtue Aristotle called *phronesis*. Thus I suggest that the general homeostatic mechanism uniting virtue is the intellectual virtue of *phronesis*, the well-functioning ability to deliberate.[18]

Though specific dependencies between virtues are plentiful and often important, I shall mainly discuss the role of *phronesis* from the HPC point of view. I will start with what Aristotle had to say about *phronesis*. Here the limitations of the source material must immediately be remarked upon. The interpretation of Aristotle's account of *phronesis* is a matter of controversy, and the discussions we have in *EN* and *EE* are insubstantial.

7. Phronesis

Interestingly, the way *phronesis* works is the very reason why Aristotle was led to think that there is something like a unity of virtues. The account is notorious, with Aristotle first explaining that to have *phronesis*, one must have the other virtues, and then claiming that being really virtuous at all requires one to have *phronesis*, for "if a man once acquires [reason] that makes a difference in action; and his state, while still like what it was, will then be excellence in the strict sense." (*NE*, 1144b11–13). That is, there are virtue-like dispositions, but without the presence of reason they can actually be harmful and lead astray. Aristotle thus draws a distinction between "natural virtue" and "virtue in the strict sense", that is involving *phronesis*. In fact, natural virtue covers more than the name suggests, as it here seems to include not only inborn natural capacities, but also what could be called proto-virtues, namely habituations not yet involving reason. Aristo-

[18] Term is variously translated as "practical reason", "wisdom" ("practical" or not) or even as "prudence". I think practical reason and prudence are safer than wisdom, but opt for leaving the word untranslated.

tle then concludes that "it is not possible to be good in the strict sense without practical wisdom, nor practically wise without moral virtue" (*NE*, 1144b31–32).

Deliberation is only of mutable matters: "no one deliberates about things that cannot be other than they are (*NE*, 1141b12). This sets it apart from *sophia*, which is concerned with eternal and immutable things, or unconditional necessities. For Aristotle *sophia* and *phronesis* are virtues of two different parts of the rational portion of the soul; *sophia*'s *ergon* is to grasp the eternal necessities, *phronesis* to deal with the mutable truths.'

Aristotle's insistence on this pattern of two-fold dependence has been much mused upon. One way to explain this peculiar standing of virtue-like capacities which are not yet virtues is to draw from Aristotle's account of *akrasia*. One can then make use of the notions of incontinence and continence and remark that virtue-like capacities can exist in isolation in an agent who is continent (having habits which need forcing, as it were) or even incontinent (having habits which tend to fail despite the agent's knowing better). But only when an agent can be said to be really virtuous, these previously separate capacities function together in a reciprocal way. (Very roughly this was the view of Henry of Ghent.) Thus the account of *akrasia* would also be of central relevance to the psychology of moral education, as continence and incontinence would be characteristic stages in agent's developing virtues.

Consider Scotus against Henry: "Accordingly, each virtue will be the ground for every other virtue's being a virtue. The consequent is false, because it follows that something is a virtue before it's a virtue, and thus no virtue will be first" (ibid., 27). According to Scotus, the following is sufficient to show that each virtue-like capacity can exist by itself. Assume two habits, which if perfected in a suitable way, would become virtues. Now, either one could be perfected without the other by performing some act or two virtues could be constituted simultaneously by one act. But then either the possibility of one habit being perfected independently is proved or some act contributes equally to the coming about of two virtues, which Scotus assumes implausible (ibid.). Extrapolating this to all virtues, the implausibility becomes apparent, for surely no one act could perfect all our dispositions. Necessity of any mutual entailment between proto-virtues, or virtue-like habits antecedent to the formation of fully functional *phronesis*, is thus denied by Scotus.

According to Aristotle virtues divide into two main categories, moral and intellectual. This division stems from his way of partitioning the soul into vegetative, animal and rational parts. Each of these parts has its characteristic functions. While the proper functioning of the vegetative part does not fall into the scope of ethics, the animal part is the proper subject of character traits, the perfection of which results in moral virtue. Intellectual virtues are perfections of the rational functions of the soul. They include capacities for theoretical knowledge and various other skills. Among the intellectual virtues, *phronesis* plays a significant role in Aristotle's account of virtues. As for the matter of the ultimate human goal, there is a persistent debate on whether Aristotle thought that the exercise of the purely theoretical capacities, that is the other half of the rational soul, has an essential role in it. Nonetheless, the other intellectual virtues do not display similar downward function with respect to moral virtues, except in so far as sound reasoning and knowledge are presupposed.

It is possible to have some, but not all of the so-called natural virtues, but this does not apply to the real deal (*NE*, 1144b30–1145a6). Such cases, according to Aristotle, are apparent in children and animals, a subject often remarked upon in *Historia Animalium*. Animals and children (again, according to Aristotle) lack reason, and hence cannot have the virtues proper for the intellective part of the soul.

In particular, *phronesis* plays a key role in being the ability to recognize the mean which is the virtue between two opposing vices. In the end, only a person of practical wisdom, a *phronimos*, is fully fit to determine this; virtue is the mean which such a person would determine (1107a2). Phronesis is such that those having it deliberate well. Aristotle tells that a *phronimos* is characterized by good deliberation (*NE*, 1141b9-12). Practical wisdom then is the ability to deliberate or to calculate, and to discover not only the means to an end, but also the mean which is so central to the Aristotelean conception of virtue.

Typically it has been thought that *phronesis* provides us with information concerning the ends as well, but this too has been challenged. A notorious formulation by Aristotle suggests that the goal is set by virtue alone, not by deliberation: "Virtue makes the goal right, practical wisdom the things leading to it" (*NE*, 1144a7–8; see Moss 2011). This sort of deliberation must be distinguished from what Aristotle calls cleverness which is a morally neutral capacity, also shared by some animals (*NE*, 1144a23–8). Cleverness can provide calculations, but such capability is not perfected in the way required of *phronimos*.

The strong Aristotelian dependency of all virtues on *phronesis* may seem to clash with Boyd's insistence on the possibility of imperfect homeostasis. But of course the fact that the notion of homeostatic clustering in some cases allows for imperfection doesn't require this to be the case always. In some cases, with fundamental physical kinds, say, it could well be the case that imperfect clustering never occurs. But *prima facie*, perfect clustering seems less plausible the further removed we are from physical or chemical kinds; already the biological world would seem to offer plenty of examples of imperfect clustering, and psychological and social spheres even more so. It should be granted, apparently contrary to Aristotle, that imperfect clustering does seem to occur with virtue.

Imperfection in the actual clustering can be dealt with by the resources here introduced, namely the distinction between types of clustering and different kinds of dependencies between virtues. For example, the apparent imperfect clustering of virtues can really turn out to be imperfect clustering of proto-virtues, by their various specific dependencies. But the Aristotelian notion of *phronesis* still appears too strong to be incorporated as such in a reasonable naturalistic moral theory. This is not caused by the patterns of dependence between virtues, or by the central role of *phronesis*, but rather is due to the biconditional version of reciprocity. This is gives rise to problems, not unlike the one from Scotus just cited.

One way out is to weaken the connection between *phronesis* and moral virtues adequately so that it conforms to our contemporary assumptions on moral psychology. Another way is to see the *phronimos* and his quality as a counterfactual paradigm case which the various imperfect agents approximate. In this view, *phronimos* is not a real individual, but an ideal model constructed from qualities abstracted and extrapolated from the observable instances and mechanisms: an ethical counterpart of a frictionless plane or an infinitely deep ocean. Nonetheless, individuals of the real world approximate this model to various degrees.

Weakening the biconditionality is preferable in any case. Let us say that the biconditionality obtains in only in the ideal case of *phronimos*, understood in the way just explained. In this theoretical sage, perfected character and intellect work without friction. But in the real world, agents fall short of this. It is nonetheless perfectly reasonable to think that in them, too, there obtains a connection whereby reason filters out the right mean with regard to action and emotion. I suggest that this relation comes in degrees, rather than as something either had or not had. By attaining a certain level of proto-virtues one begins to develop ways of properly regulating their functioning and interplay and to perceive the related means. The influence of reason back on the developing virtues should also be seen as coming in degrees of fallibility and control.

The reasons for this could be summed as follows: Aristotle's answer to the dialectical argument does not sufficiently distinguish between inborn natural virtues (such as possessed by various animals) and the learned, but not yet perfected virtues (which I have been calling proto-virtues). It seems contrary to his claim possible that one exhibits a genuine approximation to proper virtue and has a level of *phronesis*, but nonetheless the clustering of virtues remains imperfect. There is nonetheless unity, even if biconditionality fails. The homeostatic mechanism, which habituation creates gradually, provides for this. It seems plausible that as it develops further, the clustering also becomes tighter as the agent becomes better in perceiving the various means and their interconnections.

If this picture of unity this is right, inductive generalizations should be indeed expected to work well with virtues and with other types of natural goodness. For the strong varieties of the unity doctrine, this indeed becomes a trivial matter, for certainly identical or biconditional traits imply each other in any case. But we have here given a more flexible account of the unity, one in which there is room for inductive generalizations.

While such generalizations are in the end an empirical matter, let it be here remarked that this hypothetical inductive unity is up to a degree suggested by our tendency to assume a certain level of unity of virtue, and to be surprised when this assumption fails (see Hurtshouse 1999, 119). Similar tendency applies to vice, where one vice in a person is easily taken as sign of warning about further failures of moral character. Likewise, from given agent's virtuousness simpliciter, you can infer to various individual virtues of the agent. This also applies to natural goodness attributions, treated in the penultimate section.

I have mostly avoided the topic of specific dependencies. They are a mixed bag, partly *a priori*, stemming from conceptual overlaps between different capacities. In other cases they are *a posteriori*, and as such not traceable without empirical investigation. But sometimes it is hard to tell whether the dependence is of one sort or another. Quite typically, however, some preliminary insight about the connection is available for us, being made transparent by our role as acting agents. Such internal perspective on action provides us with a grasp of these conceptual connections, as we need them to mold our own functioning. But such a grasp is as fallible here as it is elsewhere.

8. The Scope of Unity

Badhwar says that some character traits, such as "[c]aution and spontaneity are obviously independent, and when highly developed, mutually incompatible" (1996, 306). But she continues to claim that this does not suffice against Aristotle on the matter of unity, as the thesis, according to her, is "meant to be true of the major Aristotelian virtues – justice, courage, temperance, generosity, and kindness" (ibid, 307). Now, I think that in Aristotle there is no indication that the scope of unity is to be limited in this manner, and that the way Aristotle insists that virtue proper is to be informed by *phronesis* in fact precludes any such limitation. This is not to say that it is not a reasonable suggestion that these central virtues contribute more to the development of *phronesis* than some peripheral ones. I am inclined to say that this question is largely an empirical one, to be for example studied in developmental psychology.

But Aristotle aside, the suggestion of limited scope is of interest. Our account indicates that there may be local clusters among virtues, such as the connection between piety and justice in Socrates's argument against Protagoras. But there are two senses of unity with regard to the relation of connected virtues, direct and indirect, and the key sense at play in Aristotle is the indirect one. This entails that the claim holds globally, even if this would only be completely realized in the ideal, counterfactual *phronimos*. This, however, does not yet show that there are no other connections. It seems plausible that some bundles of moral virtues cluster in more than one way. I have indicated above how natural virtues or developing moral virtues (proto-virtues) could form clusters by themselves. There is no reason to assume that the originally connected traits lose this clustering when they come to function under *phronesis*.

The above suggestion that the unity of virtue applies only to some central cluster among virtues is not the main claim advanced by Badhwar. Rather, in her view the unity applies in limited domains, that is, in limited spheres of action such as at home or among friends, and that in such spheres any one virtue entails the rest (1996, 308). Under this conception, having a virtue in one domain then does not entail that it is had in others, but it does entail that one does not have a corresponding vice elsewhere (ibid.). How this relates to what has been here put forth is a further issue. I will only remark that what was said in the previous section about the gradual acquisition of *phronesis* should be seen as analogical to Badhwar's limited domains. Surely, in the ideal case, no domain specificity of this kind would occur. But learning to widen the application domains of virtues too is a gradual task, and one that we should not too readily assume easy to achieve.

So where do we stand with all these distinctions, exegetics and elaborations? While the above discussion indicates that the unity of virtue, understood as homeostatic phenomena, is plausible, the issue also becomes rather complex. Many of these complexities are empirical rather than conceptual issues, and in the end I believe that the unity of virtues should be treated as an empirical hypothesis. But one of the lessons here is that the appearance that putative virtues can occur individually is misleading, and does not suffice to refute the thesis.

9. Natural Goodness approach and Clusters

In a sense, nothing that has been said so far really necessitates moral realism. Suppressing the fact that we took off from Boyd's realism, with minor alterations the account above could be read as ethically neutral classification of certain persisting psychological factors making use of the HPC conception, which by itself does not entail Boydian treatment of *Good* etc. as natural kinds.[19] But the Aristotelian picture comes with an inbuild variety of moral realism, and this shows some conflict and some resonance with the Boydian metaethical picture.

A good thing is one which does its job well. For every proper subject of evaluation, there is something it is for, in Aristotle's greek its *ergon*. And this is something it can fulfill well, in which case it is a good object of its kind. It then has virtue, or *arete*, a proper well-functioning for an object of its kind. Many virtue ethicists have sought analogical accounts to fix a central, naturalistically treatable meaning among the various uses of good.

A distinction made by Geach between attributive and predicative adjectives can be used here (Geach 1956, 33; Foot 1994, 162–164). Attributive adjectives are applied only in conjunction with a noun, and they cannot be split from the compound, unlike the predicative ones. Applied to the evaluative case, this tells us that an attributive adjective can only be used to characterize something as, say, good or bad when it occurs in combination with a kind term telling what sort of thing the thing evaluated is; that is, "a good horse", "a good fireman", "a good tennis racket" etc. These cannot be split meaningfully into two independent attributions, to get for example "this is good and this a tennis racket." Seen in this way, excellence

[19] Antirealist could just take the neutrally described psychological items marked as "virtues" and add some exclamation marks. Some like Harman of course insist that no virtues exist, but that is a different matter.

thus is essentially related to a thing's kind, and in the case of living things, their species.

Philippa Foot in particular advanced a prominent defense of such a conception (see especially Foot 2001, for critical discussion see Lenman 2005). Following Michael Thompson (Thompson 1995), Foot made use of the so-called "Aristotelian categoricals". These are unquantifiable generic statements which attribute to living thing properties characteristic of its species. This move makes evaluation depend on the species of the thing evaluated. As Thompson put it, "an appeal to notions of life and organism and life-form would seem to be implicit in all departments of ethical thought" (1995, 250).

More specifically, Aristotelian categoricals make attributions of what is normal for a member of a given species qua member of that species, what is normal in the species as it were. Goodness thus is in Foot's earlier jargon "autonomously species dependent" (Foot 1994, 163).[20] Such evaluations also set these generalities against the background of how the species lives. They are generalities which matter for the success of living beings as such beings of their kind. Hence such categoricals ground norms for evaluating a member relative to its species.

Aristotle, too, was keenly aware of this sort of species relativity. He says in *Nicomachean Ethics* that "there will be a number of different kinds of wisdom, one for each species: there cannot be a single such wisdom dealing with the good of all living things, any more than there is one art of medicine for all existing things." (*NE*, 1141a30–35). Furthermore, in *Physics* (246a13–16) we find that a thing's excellence is tied with what is natural for it: "excellence is a perfection (for when anything acquires its proper excellence we call it perfect, since it is then really in its natural state […]), while defect is a perishing of or departure from this condition." This also applies to evaluations of persons "for excellences are perfections of thing's nature and defects are departures from it" (*Phys.*, 246b1–2; as given by Hutchinson 1986).

[20] "Autonomy" here means that the goodness of the thing is not evaluated from the perspective e.g. of usefulness for human purpose. Artefacts of course cannot be subjects of such evaluations.

Another "Aristotelian" notion Foot makes use of is Anscombe's "Aristotelian necessity". This sort of necessity indeed makes its appearance in Aristotle, who calls it hypothetical necessity, and distinguishes it from absolute necessity and necessity of coercion (*PA*, 639b20-32; 642a3-9 ; Met. iv, 5).

Obviously, such an account is bound to come with a warning when applied to humans, as few could expect that our biological features could or indeed should fix what it is to be a good human, or what is defective in a human. But more of this elsewhere; another problem is that Aristotelian categoricals appear to introduce the suspicious notion of natural teleology – and no wonder.

A brief answer to this worry: Teleology hereby introduced need not be very deep. No Prime Movers are smuggled in, and no cosmological revision is intended. Rather, to put the matter in jargonistic terms for the first approximation: the purposefulness invoked here is local and can be treated as supervenient on evolved natural facts concerning the species, its means of survival and its environment.

Another related concern is that the notion of species which enters essentially into Thompson's and Foot's account is not sufficiently clear. If the notion of species itself hangs on thin air, so surely does the notion of the "life form of a species". This is an interesting point, and I hope to sketch an answer here. I believe that the answer can be found in the HPC conception of biological taxa. Another side of this coin is to see whether following this track leads to a more informed conception of the relation between functional goodness and the HPC-conception.

Boyd's HPC approach has evident suitability to biological kinds (species), and the account has been further developed by Boyd and others for that purpose (Boyd 1999, Boyd 2012). Roughly, on the HPC conception, a species is a co-occurrence of certain properties bound by a homeostatic mechanism. Further tuning is needed to accommodate for example sexual dimorphism, conditions of common descent and reproductive integration (Boyd 1999, 165, 167). While this Boydian conception preserves some features of the essentialist conception of species (such as presumably was Aristotle's), it also relaxes it considerably, doing away for instance with the assumption of definability by necessary and sufficient conditions (ibid. 1999, 145–146).

In fact, it is quite plausible to suggest that whatever way species are fundamentally treated, they have to exhibit a level of homeostatic, mechanically bound, clustering: "It is, I take it, uncontroversial that biological species, whether or not they are natural kinds, are phenomena which exhibit something like the sort of property homeostasis which defines homeostatic property cluster natural kinds." (Boyd, 1999, 164–165). We thus need not commit ourselves here to the stronger claim (defended by Boyd) that species are HPC kinds – it is sufficient for our modest purposes to observe that they nonetheless evidently manifest this sort of homeostatic clustering. But where does this observation leave us with teleology and Aristotelian categoricals?

HPCs actually do provide material for solving this. Here is a suggestion. Boyd says: "A variety of homeostatic mechanisms […] act to establish the patterns of evolutionary stasis which we recognize as manifestations of biological species" (ibid.) Now, judging natural goodness is a matter of evaluating performance with reference to some function. By examining an HPC we can tell what kind of thing a thing is, even if it is not definable in the manner Aristotle would have assumed. An HPC *sans* the mechanism is in fact not unlike a list of Aristotelian categoricals. Add the mechanism(s), and you have a nice approximation to what is the *telos*.

What *Arete*-goodness attributes of subject is its functioning well, doing its job well etc. Now there is a seeming discrepancy between the Boydian cluster view and this sort areteist concept of goodness, because HPCs are primarily made of clustering properties as such, and areteist goodness is not a simple property of this kind but rather a property of fulfilling a function in a certain way. By dissolving this issue by putting function and cluster in their right places, we also avoid Rubin's arguments in so far as they would pertain to our case.

In so far as functions are properties of first order, areteist good is then of a higher order, in that it characterizes workings of certain other properties. But in another sense it is not higher-order, because it also characterizes its subjects as such. For example: a knife is a good knife, if it does its job well;

knife's nature is such that it ought to cut well. Hence, its kind (which is a property) contains a function, a reference to something it is for. And in performing well with respect to this it becomes called good.

At the bottom of this is nothing over and above the physical constitution of the thing. A knife's sharpness (its virtue) stems solely from its material arrangement. Now, in things with complex constitutions more factors operate, and they can aid the performance of the overall functioning of the thing or hinder it. Moreover, such constituents have causal relations with each other, sometimes interfering harmfully with the overall function.

I believe that these considerations together with Rubin's arguments make Boyd's attempt to treat Goodness as natural kind in the consequentialist manner problematic. Nonetheless, the natural goodness account turns out to save many of the better sustained features of the Boydian outlook.

10. Conclusion

Virtues form a plausible candidate for an HPC. This sets a novel framework for a generally naturalistic account of virtue. The Aristotelian scheme of dependences among virtues provides a model on which such a theory can be built. In particular, the intellect can be seen as providing a central homeostatic mechanism. Understood in this unorthodox way the ancient idea of unity of virtue becomes something of a plausible hypothesis.

A species relative natural goodness account works better as a naturalistic HPC conception than Boyd's consequentialist picture of goodness. It integrates particularly well with biological grounding forming a continuum of evaluative notions of well-being and well-functioning. Rubin's argument against Boyd partially miss their target due to Rubin's error in characterizing Boyd's position. Nonetheless, a natural goodness account is more successful against these arguments, lending more support for the position.

Rubin's arguments for structural or inductive failures are avoided on the natural goodness approach to HPC moral realism. Goodness itself is not treated as a cluster concept, but it is closely associated with the cluster which is the HPC kind of its species. The properties that constitute virtue are such that Rubin's objections do not rise against them. Each virtue is such that on its basis Goodness *qua* good performance of a given function can be predicated of its bearer. But concerning natural goodness *simpliciter* is not so predicable on this basis, because such goodness is not attributable in the absence of most of the properties in the cluster, as is to be expected. Goodness is a matter of unobstructed, non-deficient functioning of the various parts of the homeostatic mechanism. In particular, inductive generalizations work well with virtues and vices. Many paradigmatic examples of clustering witness this.

Virtues and their failures cluster, and so having some makes it more likely that one also has others. The biconditional Aristotelian conception would turn this propensity into a necessity. But we can drop that while keeping a good part of what is right in the theory.

University of Helsinki

References

Adams, Robert Merrihew (1999), *Finite and Infinite Goods. A Framework for Ethics,* Oxford University Press, Oxford.

Adams, Robert Merrihew (2006), *A Theory of Virtue. Excellence in Being for the Good*, Oxford University Press, Oxford.

Annas, Julia (2011), *Intelligent Virtue*, Oxford University Press, Oxford.

Aristotle, *Categories* (translated by J. L. Ackrill) in Barnes (ed.) (1984).

Aristotle, *Nicomachean Ethics* (translated by W. D. Ross) in Barnes (ed.) (1984).

Aristotle, *Eudemian Ethics,* Translated by Brad Inwood and Raphael Woolf. Cambridge University Press, Cambrige. (2013).

Aristotle, Metaphysics, (translated by W. D. Ross) in Barnes (ed.) (1984).

Aristotle, *Parts of Animals*, (translated by W. Ogle) in Barnes (ed.) (1984).

Aristotle, *Physics*, (translated by R. P. Hardie and R. K. Gaye) in Barnes (ed.) (1984).

Barnes, Jonathan (ed.) (1984), *Complete Works of Aristotle. Volumes I–II.* Princeton University Press.

Badhwar, Neera K. (1996), "The Limited Unity of Virtue". *Nous* 30:3, pp. 306–329.

Boyd, Richard (1988), "How to be a Moral Realist?" in Sayre-McCord (1988), pp. 181–228.

Boyd, Richard (2003a), "Finite beings, finite goods: The semantics, metaphysics and ethics of naturalist consequentialism, part I", *Philosophy and Phenomenological Research*, 66, 3, pp. 505–553.

Boyd, Richard (2003b), "Finite beings, finite goods: The semantics, metaphysics and ethics of naturalist consequentialism, part II", *Philosophy and Phenomenological Research*, 67, 1, pp. 24–47.

Boyd, Richard (1999), "Homeostasis, Species, and Higher Taxa", in Wilson (ed.) 1999, pp. 141–185.

Boyd, Richard (2010), "Higher Taxa, and Monophyly", *Philosophy of Science*, 77, 5, pp. 686-701.

Cooper, John M. (1998), "The Unity of Virtue" in Cooper (1999), pp. 76–117.

Cooper, John M. (1999), *Reason and Emotion. Essays on Ancient Moral Psychology and Ethical Theory*, Princeton University Press, Princeton NJ.

Driver, Julia (2001), *Uneasy Virtue*, Cambridge University Press, Cambridge.

Duns Scotus, John (2017), *Selected Writings on Ethics*. Edited and translated by Thomas Williams. Oxford University Press, Oxford.

Foot, Philippa (1978), "Virtues and Vices" in Foot 2002 (1978), pp. 1–18.

Foot, Philippa (1983), "Moral Realism and Moral Dilemma" in Foot (2002), pp. 37–58.

Foot, Philippa (1994), "Rationality and Virtue" in Foot 2002, pp. 159–174.

Foot, Philippa (2001), *Natural Goodness*, Oxford University Press, Oxford

Foot, Philippa (2002/[1978]), *Virtues and Vices,* Oxford University Press, Oxford.

Foot, Philippa (2002), *Moral Dilemmas*, Oxford University Press, Oxford.

Geach, Peter (1956), "Good and Evil", *Analysis*, 17, pp. 33–42.

Geach, Peter (1977), *The Virtues, The Stanton Lectures 1973–74*. Cambridge University Press, Cambridge.

Hurtshouse, Rosalind, Lawrence, Gavin & Quinn, Warren (eds.) (1995), *Virtues and Reasons. Philippa Foot and Moral Theory. Essays in Honour of Philippa Foot*. Oxford University Press, Oxford.

Hurtshouse, Rosalind (1999), *On Virtue Ethics*, Oxford University Press, Oxford.

Hutchinson, D. S. (1986), *The Virtues of Aristotle*, Routledge Kegan & Paul, London.

Kenny, Anthony (1978), *The Aristotelian Ethics. A Study of the Relationship between the Eudemian and Nicomachean Ethics of Aristotle,* Oxford University Press, Oxford.

Lenman, James (2005), "The Saucer of mud, The Kudzu vine and the uxorious cheetah: Against neo-Aristotelian naturalism in metaethics", *European Journal of Analytic Philosophy,* 1, 2, pp. 37–50.

Moss, Jessica (2011), ""Virtue Makes the Goal Right": Virtue and Phronesis in Aristotle's Ethics", *Phronesis*, 56, 3, pp. 204–261.

Rubin, Mark (2008), "Is Goodness a Homeostatic Property Cluster?", *Ethics*, 118, 3, pp. 496–528.

Sayre-McCord, Geoffrey (1988), *Essays on Moral Realism*, Cornell University Press, Ithaca and London.

Sreenivasan, Gopal (2009), "Distunity of Virtue", *Journal of Ethics*, 13, 2–3, pp. 195–212.

Thompson, Michael (1995), "The Representation of Life", in Hurtshouse et al. (eds.) (1995), pp. 247–296.

Wilson, Robert A. (ed), 1999, *Species. New Interdisciplinary Essays*. MIT Press, Cambridge MA.

Vlastos, Gregory (1972), "The Unity of the Virtues in the "Protagoras"". *Review of Metaphysics* 25, 3, pp. 415–458.

Four Reasonable, Self-Justifying Values – How to Identify Empirically Universal Values Compatible with Pragmatist Subjectivism

FRANK MARTELA

Introduction

What is ultimately worth striving for in a human life? Humans typically have many projects and goals in life, but when they start to ask the 'Why?' question as regards these projects, what could satisfactorily answer this question? In other words, what are the self-justifying values that are not derivative of or dependent on other values, but provide their own justification? For the Russian author Leo Tolstoy (2000, 12) asking the 'Why?' question too many times led to an existential crisis: Attending to his estate would lead to his fields producing more crops, but what then? Unable to find anything to ultimately justify his activities, he felt that "what I was standing on had given way, that I had no foundation to stand on, that which I lived by no longer existed, and that I had nothing to live by."

Tolstoy was unfortunate to live in a period of time when the scientific worldview had unstabilized the traditional worldview where the world contained self-evident values that gave purpose to human living. A void was unveiled, and thinkers from Thomas Carlyle and Fyodor Dostoevsky to Søren Kierkegaard, Arthur Schopenhauer, and Friedrich Nietzsche were staring at it. The question was, and still is: How to reconcile a scientific worldview that seems to leave no room for self-evidently objective values with a human yearning to have something solid to base one's life decisions and goals on?

In this article I am not making an argument for the scientific worldview or its incompatibility with the existence of objective values. Instead, I take these as a starting point. Let us thus assume that values are not something found from the structures of the world. “Before life began, nothing was valuable”, as Street (2006, 155) puts it. And while naturalistic objectivism has recently become relatively popular among philosophers of value and meaningfulness (see Kauppinen, 2016; Metz, 2013), let us assume that no such objectivistic option is available for us. Instead, let us assume that values are a human construction, something we have created in order to navigate our lives, having no justification beyond what we happen to prefer and value (Dewey, 1932; Martela, 2017b). Given such subjectivism about values, is there any way to argue that one value is better than another? And is there anything that could provide a satisfactory answer to the ’Why’ question?

I argue there is. But this justification available to us doesn’t depend on anything external to us or on anything objective. Given that humans are a certain type of species, programmed by evolution to seek certain types of experiences, it is possible to identify empirical generalizations about natural human motivational dispositions. And based on those generalizations we can make recommendations that upholding certain values is *better* for the average person than upholding some other values. Better not in an ultimate sense, but better in the same sense as in the medical art (Dewey, 1939, 21). Some diets provide more of what our body needs in order to function well, and empirical research has been able to provide relatively reliable guidelines on this matter. In the same sense, some strivings and values are better aligned with what humans need to function well psychologically, and empirical research can provide reliable guidelines on this matter as well. In particular, we can identify what humans are naturally prone to seek. In other words, empirical research can help us identify the things we are typically motivated to pursuit as a species because of certain dispositions acquired through evolution. This information can then be used to make the case for values that are well aligned with these dispositions. The values identified through this method would not be objective in any sense, but they would have three qualities that would make such values feel worth committing to, even after careful reflection. First, if a potential explicitly upheld value is well aligned with such a natural motivational disposition, this can make the value *feel* like its own justification. Also, explicit values closely connected with these implicit motivational dispositions would have a relatively universal motivational pull across cultural boundaries and, finally, such values would be closely connected with good psychological functioning and wellness, making them especially attractive candidates for what to reflectively value. The general aim of this paper is thus to suggest one mode of inquiry through which to identify self-justifying values that many different subjects would find reflectively worth valuing. This is less than what objectivists would like to have, but in a world where the quest for ultimate certainty is prone to leave one empty-handed (see Dewey, 1929) this might be the best option available to us as regards reliable and warranted guidance on what to value.

In what follows, I first make a distinction between worthy and meaningful lives, and between derived and self-justifying values, in order to set the stage for what we are searching for. Then I offer my suggestion about what characteristics a subjectivistic yet empirically warranted self-justificatory value should have. Furthermore, I identify four prime candidates for such self-justifying values and for each of these potential self-justifying value, briefly discuss how they could be grounded in what we know about the human nature. I conclude by suggesting that these four self-justifying values could be seen as our most empirically-grounded generalization about what could make a life worth living in a silent universe devoid of objective values.

Distinctions: Worthy and meaningful lives, derived and self-justifying values

When we search for self-justifying values for life, what are we actually searching for? *Meaningful* lives on the one hand and *worthwhile* or *worthy* lives on the other hand are often used as synonyms for each other, but I see that a distinction between them could clarify our search. Meaningfulness of a life is a certain type of evaluation we can make about life, in the same sense that we can evaluate the happiness or pleasurableness of that life (Wolf, 1997). There is considerable debate, into which we don't need to go here, about what exactly meaningfulness as an evaluation is about (see Metz, 2013). But most agree that it is an evaluation about a certain type of value a life can have that is not reducible to other values such as happiness. However, beyond these evaluations about specific types of values a life can have, we need a label for the overall worthiness of a life. Such an evaluation of worthiness of a life "takes into account all possible things that can influence the judgment whether a certain life is worth living and whether a certain life is more choiceworthy than another life" (Martela, 2017a, pp. 234–235)[1]. It is the broadest evaluative question we can ask about a life, similar to what Haybron (2008, 36) calls a good life: "a life that is desirable and choiceworthy on the whole: not just morally good, or good for the individual leading it, but good, all things considered—good, period."[2]

There are several things affecting such an overall judgment of the worthiness of a life. Moral goodness, meaningfulness, and happiness have been already mentioned. Many things making a life good are reducible to these values. Most would agree that having food on the table and being safe from predators such as wolves make, other things being equal, a life more worth living. But these things are not good as such, but good because they represent the absence of certain worries and sources of unhappiness in one's life and thus can signifi-

[1] Here I use good, worthy and worthwhile lives as synonyms for each other. If one would want to make a further distinction between worthy and worthwhile lives, one could say that the latter denotes a life that is worthy enough to pass a certain threshold and thus be "worth the while".

[2] It is worth noting that this way of evaluating the worthiness of a life is essentially a subjectivist evaluation of the choiceworthiness of a certain life path and way of living over another. This is a different notion than a moral or political evaluation of the value of a certain human being. Human lives have a certain intrinsic dignity as such. This seems like a good premise for political debates as we should not easily start evaluating one life as more valuable than other when making decisions as political actors. But as living beings making choices about our own course of lives, we inescapably rely on a more or less implicit notion of what factors influence the choiceworthiness and goodness of a life, which I here aim to make more explicit.

cantly influence the happiness of the person living a particular life. Their goodness is thus reducible to happiness, they derive their value from it. Thus the value attached to such a thing is instrumental or derived, it emanates from some more basic value they contribute to. But pleasure or happiness (with which I mean psychological happiness, Haybron, 2000) needs no justification beyond itself, but seems to be something humans typically value on its own accord (e.g. Mill, 1863). *Self-justifying value* is the term for any value that a person readily acknowledges as valuable on its own accord, that is not reducible to or derivative of other values, and that thus is by itself its own justification.

Self-justifying values need not be objective in any sense, self-justifying only implies that the subject in question sees no need to seek further justification for the value in question. The criterion of what makes a certain value self-justifying is thus subjective: The value is self-justifying to the degree that the subject experiences it as requiring no further justification. More generally, when I talk about values in this article, I am not making references to anything objective or human-independent. Following Sharon Street (2006, 118), I see that "the capacity for full-fledged evaluative judgments was a relatively late evolutionary add-on, superimposed on top of much more basic behavioral and motivational tendencies." I see no ontological or epistemological gap between mundane everyday desires and preferences on the one hand, and more explicitly-held values on the other hand. The difference is found merely in the degree of abstraction and in the degree of conscious commitment (Dewey, 1938; Martela, 2015). The kind of values that we seek here are thus nothing more than motivational preferences that we have reasoned that we want to reflectively endorse, commit to, and hold in so high regard that we are willing to base our major life decisions on them[3].

[3] This explains how the present account aims to avoid the 'naturalistic fallacy' (Moore, 1903) of leaping from what people actually value to what ought to be valued. Here, I aim not to make any claims about what is ultimately worth valuing or what *ought* to be valued. I am merely claiming that humans tend to value several things, and for some of the things we value we can find more reflective reasons to value them. And here I aim to offer a few such reflective reasons that we can use when judging whether or not to value certain things.

These self-standing sources of value are then the independent dimensions we use in the evaluation of what makes a life worthy and good. And there seems to be a relatively limited number of strong candidates for such self-justifying values. This narrows down our search: When asking what is worth striving for in human lives, we are searching for the key self-justifying values that could offer guidance for human living.

But how to identify and evaluate the potential candidates for self-justifying values?

Empirical universalism: A pragmatist naturalistic account of self-justifying values

Philosophers have tried to identify the intrinsic and self-justificatory goods that humans strive for at least since Aristotle's (2012, book 1, chapter 2, 19) famous search for "some end of our actions that we wish for on account of itself, the rest being things we wish for on account of this end."

In arguing for and against potential intrinsic values, modern philosophers, especially of the analytic bend, typically examine whether the theory can explain the most intuitive cases of value, and not fall prey to various counterintuitive conclusions. In other words, they rely more or less on intuition (see Metz, 2013, 8). For example, discussions about meaningfulness as an intrinsic value typically concentrate on formulating theories of its nature that can explain the prototypical examples of meaningfulness and steer clear from the commonly accepted counterexamples (e.g. Martela, 2017a; Smuts, 2013). But the fact that meaningfulness intuitively sounds as something valuable and worth striving for is usually taken as given.

Here, instead of reasoning to explain our intuitions, I suggest to complement it with a more empirical strategy for identifying self-justifying values. More particularly, I see that humans, as well as any other species, have acquired through evolution certain motivational or proto-motivational mechanisms. Beyond the *explicitly held values* that humans are consciously aware of and committed to, there are *implicitly held preferences* or proto-values that guide our behavior and thinking even when we are not aware of them (Haidt, 2001; Street, 2006). Simply put, we are naturally drawn towards certain things and naturally aversive of other things. A bacteria navigates towards certain chemicals, a flower reaches towards the light, a lion seeks water to drink and herbivores to eat. Certain key resources such as air, water, and food are crucial for animal survival, and thus also human psychology includes motivational mechanisms that have developed to ensure that the organism behaves in ways that typically lead to the acquisition of these key resources. However, a social animal like human being designed to live in relatively large tribes (Dunbar, 1998) have developed motivational dispositions that go beyond mere acquisition of food and water. The survival and reproduction opportunities of a human individual have been highly dependent on one's position and reputation within one's tribe. Accordingly, the rapid threefold increase in hominid brain size taking place in the last 2 million years has been described as a within-species arms race of increased social skill to handle and keep track of social collaborations and competitions (Bailey & Geary, 2009; Flinn, Geary, & Ward, 2005). This has presumably also led to a more complex pattern of basic motivational dispositions, defined here as evolutionary acquired natural motivational tendencies to seek certain psychosocial experiences, especially when such experiences are lacking in one's life (Martela, 2018). For example, there is quite wide consensus among psychologists that humans have a basic psychological need to experience relatedness and belonging in the sense of feeling that there are people one cares about and who care about oneself in one's life (Baumeister & Leary, 1995; Deci & Ryan, 2000). Given the existence of such basic motivational dispositions, humans thus have certain experiences that they very naturally and intuitively seek. Although the ultimate explanation for the existence of these dispositions is evolutionary fitness, on a proximal and phenomenological level the pull of these dispositions will feel intuitive to the subject; they feel they are seeking these experiences on their own accord (Ryan & Hawley, 2017). For example, we humans care about our children not because we consciously calculate how their survival increases the chance of our genes to spread, but because we mammals have a natural motivational tendency to love and care for our offspring (Marsh, 2016).

Now, if we would have a consciously upheld value that would be closely aligned with such a basic motivational disposition, this disposition would provide a strong and robust intuitive motivational appeal for this value. Let's say a person consciously decides to start upholding friendliness as a value. If humans would have a motivational disposition that would make friendly behavior motivationally appealing to them, then the behaviors recommend by the explicit value of friendliness would be supported by them feeling intuitively appealing. Experiencing something as intuitively highly appealing makes it *feel* valuable. And such a feeling would provide an intuitive justification for the explicit value. The value would be experienced as self-justificatory, because we are designed by evolution to feel a motivational pull towards behaviors recommended by the value. In other words, this value in question would exhibit the key characteristics of a self-justificatory value, a value that provides its own justification.

In addition to feeling like its own justification, such explicit values building on basic motivational dispositions would have two additional qualities that would make them attractive also on a reflective level. First, given that basic motivational dispositions, due to their evolutionary nature, recommend experiences that tend to be good for the organism and its physical and psychological wellness and functioning (Baumeister & Leary, 1995; Deci & Ryan, 2000), orienting oneself towards these experiences through upholding them as values and goals also tends to lead to increased well-being (e.g. Niemiec, Ryan, & Deci, 2009). Thus such values would be good from the point of view of wellness and psychological functioning of the organism. Second, given that such motivational dispositions would be present across cultures due to them being part of the human nature, values built upon them could offer some cross-cultural common ground upon which an agreement about the basic cross-cultural values could be built.

Accordingly, my suggestion is that, in identifying and evaluating self-justifying values, one promising strategy would be to examine how well the potential self-justifying value is aligned with human basic motivational tendencies.

Instead of appealing to mere intuition in arguing for a certain self-justificatory value, we could thus examine how strong case can be made for a corresponding basic motivational disposition, given the current psychological and evolutionary research behind such a disposition. If such a case can be made, this would provide evidence that a corresponding self-justificatory value would not only feel intuitively appealing to the investigator in question, but would have broad and robust appeal across cultures and societies. The self-justificatory values identified through this method would thus have a motivational justification that is *empirically universal*. Search for such empirically universal values aligns well with the Deweyan pragmatist philosophical tradition, where it is seen that a moral theory must be "based upon realities of human nature" (Dewey, 1922, 11) and where there is an attempt to ensure that our natural emotions, desires and needs are integrated with more conscious and culturally-based ideas and appraisals (Dewey, 1939, 65).

Empirical universalism as regards a value means that almost all members of the human species would feel its pull. This allows for some exceptions. There could be individuals or even whole groups for whom a certain self-justifying value would have no motivational pull due to some brain abnormalities, developmental disturbances or due to an upbringing that has actively sought to uproot this disposition. But for most people in most cultures, these self-justifying values would have a natural appeal. Analogously, the fact that humans depend on eyesight for navigating the world is an empirically universal assertion about the human species, and something we design our societies around. Yet there are people who by birth or through some accident or sickness lose their sight, and special accommodation has to be made to help them survive in this society built on the premise of eyesight. Empirical universalism thus aims to identify those factors of humanity that most members of the human species share, accepting a small degree of exceptional individuals and groups.

If the values we have depend on humans valuing them and the possibility of objective values is foreclosed (as was the premise of this article), this empirical universalism is arguably as close to universalism we can get as regards self-justificatory values. The empirical way of identifying self-justifying values suggested here thus could provide as robust and as warranted self-justifying values as is possible, if we make the identification of human values into an empirical science.

Happiness as a self-justificatory value

What would then be the self-justifying values that would pass the empirical test outlined above? The most self-evident of the self-justifying values in the Western philosophical tradition is arguably happiness or experienced well-being. Humans seek pleasure and avoid pain, as the hedonistic theories hold (Wolf, 1997). Not as a means to something else, but because pleasure and avoidance of pain are good in themselves. This basic truth about human nature has been recognized both by Aristotle (2012) and by John Stuart Mill (1863), although especially the former emphasized that this is not the only thing that humans seek. I use the term *happiness* to denote human inner states – feelings, emotions, affects, and so forth – that feel good rather than bad. Other things being equal, humans prefer positively valenced inner states to negatively valenced inner states. And this preference doesn't seem to need any further justification. Eating an ice-cream is enjoyable as such. We don't need any other reasons for it beyond the pleasure we derive out of tasting this sweet delicacy. Often it is taken as so self-evident that humans seek pleasure and avoid pain, that many philosophers, economists and other thinkers have felt the urge to reduce all human motivations into this single factor (e.g., Bentham, 1789).

Evolutionary speaking, the fact that humans are motivated to seek positively valenced inner states is easy to explain. The whole existence of an animal capability to experience certain states as positive or negative is designed to guide our behavior. Physical pain is a signal system that alerts us to threats to our physical body. Tissue damage such as a wound feels painful and aversive precisely because it is crucial for animals to avoid tissue damage in order to survive. Pleasure-bringing things such as fat, sugar, and sex bring such great amounts of joy precisely because such a positive feedback leads us to seek them more, and seeking them has increased the survival and reproduction chances of our ancestors.

Thus it is easy to accept that one key dimension we examine when evaluating the worthiness of a life is how much pleasure and how much pain there is in that life for the person living that life. Other things being equal, we would prefer a life of less pain and more pleasure. Humans might develop complicated relations to pains – having certain pains might be mixed with pleasure as any sado-masochist or triathlete knows, and not being able to avoid certain pain might lead one to accept and even endorse it as a defense mechanism. But in general it remains a robust fact that people are motivated to avoid pains and seek pleasures. Thus happiness can be seen as one empirically universal self-justificatory value that people across the world take into account in making choices about their lives and in evaluating the goodness and worthiness of certain lives and periods of lives.

Morality as a self-justifying value

In evaluating the overall goodness of a life, many would argue that mere focus on happiness is not enough (e.g. Haybron, 2008; Wolf, 2016). A happy but morally base life doesn't sound too attractive. If the price of our personal happiness is grave wrongdoings against others, many would not be willing to pay it. A recent news story told about a woman and her husband's best friend, who decided to kill the husband in order to live together happily ever after. Their crime remained a mystery for more than 20 years, before new DNA tests finally revealed the truth. But even if we would assume that they would have truly lived happily ever after, never getting caught, many would still find such a life bad and less choiceworthy compared to a life where happiness would be attained without having to murder anybody.

We seem to care about morality and the moral goodness of a life. This is a separate dimension to evaluate the goodness and worthiness of a life, as many philosophers have argued (Haybron, 2008; Wolf, 2016). As Haybron and Wolf among others have argued, we typically take at least happiness and moral goodness into account when evaluating life options. Not many would choose the life of a happy serial killer. On the other hand, although we admire figures like Mother Theresa, not many are actually willing to sacrifice their own personal happiness in order to live a life that is exceptionally good morally speaking.

Again, it is easy to find support for this self-justifying value from empirical psychology and from evolutionary theorizing. Morality is ubiquitous in human societies, and even many of our close primate cousins have been shown to exhibit elements of proto-morality (de Waal, 2009; Warneken & Tomasello, 2015). Empirically, it has been shown that human moral judgments have a strong emotional and intuitive foundation (see, e.g., Haidt, 2001; Prinz, 2008), and the same key dimensions seem to be behind our moral judgments across the world (Graham et al., 2013; Janoff-Bulman & Carnes, 2013; Shweder, Much, Mahapatra, & Park, 1997), even though different individuals and groups might interpret and weight them differently. No matter the cultural context, humans seem to care about the harm caused to others, and about fair distribution of resources (e.g., Graham et al., 2011). Accordingly, cross-cultural research in small tribes around the world has shown that nowhere do people behave according to the economic model of maximizing personal utility (Henrich et al., 2005, 2001). Also neurological research about the functioning of human sense of care and of fairness is burgeoning (e.g. De Quervain et al., 2004; Harbaugh, Mayr, & Burghart, 2007), and evolutionary explanations of the fitness benefits of such dispositions have gained wide support (e.g., Boyd, Gintis, & Bowles, 2010; Fehr & Fischbacher, 2003; West, El Mouden, & Gardner, 2011). Thus it seems easy to conclude that caring about moral rightness and goodness is an empirically universal assertion about the motivational dispositions of the human species. Thus being morally good can be taken as one of the empirically well-founded candidates for a self-justifying value.

Contribution as a self-justifying value

While happiness and moral goodness are relatively uncontroversial candidates for self-justifying values, they don't necessarily exhaust the possible dimensions people use when evaluating the goodness of a life. A third candidate self-justifying value is contribution, defined as “the positive contribution beyond itself that this particular life is able to make” (Martela, 2017a, 232). It’s thus about the impact of one’s life: What difference does one’s existence make to the wider world? Several philosophers have argued that meaningfulness is a separate self-justifying value from happiness and morality (Metz, 2013; Wolf, 2016). I see that it is this sense of contribution that we are primarily seeking, when we evaluate whether a life has meaningfulness beyond mere happiness (Martela, 2017a). The prototypical examples of meaningful lives – Martin Luther King, Mother Teresa, Abraham Lincoln, Mahatma Gandhi, Nelson Mandela, or Marie Curie – are exceptional precisely because these people had larger than life positive influence on the world beyond themselves.

As regards the separateness of contribution and happiness, let’s imagine a blob who spends one’s life watching sitcoms and drinking beer alone (example from Metz, 2012 who attributes it to an unpublished paper by Wolf) – and is completely happy with this lifestyle. The blob might be high on happiness and do nothing morally wrong, but still this is usually regarded as a paradigmatic example of a meaningless existence. Also other paradigmatic examples of meaningless lives – lining up balls of torn newspapers in neat rows (Cottingham, 2003), counting the blades of grass on Harvard Yard (Smuts, 2013), maintaining 3,732 hairs on one’s head (Taylor, 1991), or enjoying endless pleasures in an experience machine (Nozick, 1974) – could be imagined as happy (assuming a slightly obsessive passion in some cases), but what they seem to be lacking is any positive contribution beyond oneself. Accordingly, most normal observers would want to avoid such lives no matter what level of happiness those lives would promise, and no matter that they seem to involve nothing morally blameful. What they seem to lack is something beyond happiness and moral goodness: They don’t seem to matter beyond themselves, they lack any positive impact.

The separateness of happiness and contribution beyond oneself as self-justifying values is thus easy to see, but are moral goodness and positive contribution really different self-justifying values? As an example of the difference between morality and meaning as contribution, May (2015, pp.

117–119 example slightly modified) asks us to consider a rock star whose music touches and provides uplifting experiences for millions, but who is totally narcissistic to the degree that "those around him do not show up on his moral radar", and accordingly makes life miserable for all people close to him. Even though we might think that he is able to make a meaningful contribution to the world, we might have quite strong reservations about the moral goodness of his life. May also notes another significant difference between morality and meaningfulness: We usually see that people ought to adhere to the moral standards but "it is nobody's obligation to live meaningfully" (May, 2015, 137).

As another example, consider the case of Nelson Mandela, who is often included in lists of prototypical examples of meaningful lives (e.g. Metz, 2012, 437). His leadership and policy of forgiveness has elevated him into the status of one of the most admirable political leaders of the 20th century – and his example also helped South Africa to evade a civil war and the ensuing blood bath. No doubt that such a life is ever so meaningful. However, he himself acknowledged that his public role led him to ignore his family. He went through two divorces, and this is how he describes in his autobiography his children's reaction to him getting out of prison (Mandela, 1995, 600): "We thought we had a father and one day he'd come back. But to our dismay, our father came back and he left us alone because he has now become the father of the nation." From the point of view of morality, we can blame him for ignoring his moral duty towards his wives and his children. On the other hand, if he would have spent more time with his family, he could not have fulfilled his moral duty towards the nation by being the leader that South Africa needed at that time. There are thus two competing moral duties and people might have different intuitions about which moral duty is more binding, that towards the family or that towards the nation. But it is clear that in the latter case his impact beyond himself was tremendously larger, and this makes it more straightforward to conclude that in terms of contribution, the latter life was superior.

If this seems controversial, we can look at the situation from another direction: Think of two men getting released from the prison: Mandela, the father of a nation, goes on to become the admirable political leader we know. Nelson, the responsible father, declares that he has a duty towards his family and wants to finally have a chance to be present in his children's lives. Thus he takes on a smaller responsibility within the political party that allows him to be home and spend plenty of time with his beloved family. Did one of them do something that is morally wrong? In comparing the lives of Nelson, the responsible father, and Mandela, the father of a nation, we can find moral merit in both, and different people might have different opinions about whose life is more morally good or about whether one of the persons did a morally blameworthy choice. There is thus no clear consensus about which live is better from a moral point of view. However, as regards the meaningfulness and the societal impact of their respective lives, it is clear that Mandela operated on a totally different scale. Nelson's life was by no means meaningless! Taking care of one's children and helping them grow up is one of the regular sources of meaning in life. However, Mandela affected the lives of millions of people, and his example will serve as an inspiration for generations to come. Mandela's contribution was thus on a totally different level of magnitude compared to Nelson's. I hope that these examples are enough to show that the evaluation of the moral goodness of a life and the evaluation of the positive contribution of a life should be seen as two distinct evaluations even though many acts can satisfy both.[4]

As regards the empirical case for a basic motivational disposition behind such value for contribution, I've reviewed the evidence elsewhere (Martela, 2018). Even in anonymous situations where no personal benefit can be expected, people are willing to sacrifice some of their own resources to give something to others (see Engel, 2011 for a meta-analysis of 616 experiments). Interestingly, even in situations where one's donations would be crowded out dollar-by-dollar by the experimenter's donation (meaning that the recipient would al-

[4] Another difference is suggested by Metz (2013, 68): One can have a moral duty towards helping a drowning child, and the moral worthiness of one's act is not lessened even if one ultimately fails. However, "the value of help with respect to *meaning in life* is at least partly contingent on its results."

ways get the same amount, no matter what one gives), most people decide to make a donation (Crumpler & Grossman, 2008). People thus not only care about the other being helped, but want to themselves be the one's making the difference (see also Luccasen & Grossman, 2017). Furthermore, when people are able to make a positive impact, this tends to increase their own sense of happiness (Dunn, Aknin, & Norton, 2008) and meaningfulness (Martela & Ryan, 2016a), and such positive emotional effects have been replicated across the world (e.g. Aknin et al., 2013)[5]. Finally, some studies show that we tend to want to make a contribution even when it is not good from a moral point of view (Batson et al., 1999; Batson, Klein, Highberger, & Shaw, 1995; see also Decety & Cowell, 2015).

Evolutionary speaking, this tendency to want to make an impact has been explained by *costly signaling*, where altruistic behavior that is costly in the short run works as a signal to other group members, giving the member a certain respect that pays off in the long term (Gintis, Smith, & Bowles, 2001; Hardy & Van Vugt, 2006). Empirical research has indeed demonstrated how such costly altruism can bring reputational benefits that lead the person receiving more resources from others in the long run (Nowak & Sigmund, 2005; Wedekind & Milinski, 2000). Furthermore, such costly signaling might not only make the person more attractive collaboration partner but also more attractive mating partner (Jensen-Campbell, Graziano, & West, 1995), which also brings obvious fitness benefits. Thus – although a full review of the empirical evidence would require much more extensive treatment than what is possible here – a reasonable set of empirical research supports the notion that contribution as a self-justifying value could be grounded in an empirically universal basic motivational disposition to want to make a positive impact.

[5] However, many people seem to be unaware of these positive effects (e.g. Dunn et al. 2008) and experimental research has shown that people engage in prosocial behavior even when the motivation to gain empathic joy is controlled for (Batson et al., 1991). People thus seem to find value in helping others that is not reducible to mere instrumental motivation to gain positive emotional experiences from such helping.

Authenticity as a self-justifying value

The final self-justifying value suggested here is authenticity, which is roughly about ″being true to oneself, living authentically, being able to make autonomous choices and being able to express who one really is in one's words and actions″ (Martela, 2017a, 245). There is a certain intrinsic dignity present in situations where one stays true to oneself even when there is pressure to compromise in order to avoid certain harms or punishments. For example, Becker (1992, 20) argued that "autonomous human lives have a dignity that is immeasurable, incommensurable, infinite, beyond price." Similarly, existentialists (e.g. Kierkegaard, 1992; Sartre, 2007) and humanistic psychologists (e.g. Maslow, 1968; Rogers, 1961) typically promote the value of not yielding to external pressure but daring to live authentically and true to oneself.

That authenticity is different from contribution as a self-justifying value is relatively easy to see. One is about remaining true *to oneself*, the other is about contribution *beyond oneself*. Although there are activities where one is able to fulfill both values simultaneously (e.g., a nurse who truly enjoys her work activities), they can easily come into conflict. William Damon (2008, 20) gives the example of a cardiologist whose surgical skills saved human lives on a regular basis (contribution) but who "hated his work to such a degree that he could barely get out of bed in the morning" as he felt that he had chosen this career just "to please other people" (inauthenticity).

Also the difference between authenticity and morality as self-justifying values is relatively clear, with philosophers like Nietzsche (e.g., 1961) emphasizing their separateness. Expressing oneself and remaining loyal to moral standards can sometimes be in conflict with each other. A dedicated sadist might choose to not live out all of one's fantasies, as some of them might be unjustifiable from a moral point of view. An autobiographical author such as Karl Ove Knausgård might expose much more about the private lives and secrets of those close to him than what would be morally good, in order to stay true to one's inner artistic vision. There are situations in most lives where one must balance the desire to express oneself and remain true to who one truly is with how much one is willing to deviate from the morally good behavior.

The difference between authenticity and happiness as values might be less clear. Both seem to refer to something within the individual: remaining true to oneself and promoting positive inner states. Empirical research has also demonstrated that a sense of autonomy is an important predictor of experienced well-being (see Deci & Ryan, 2000). Accordingly, one could argue that authenticity is only a means to happiness rather than a separate self-justifying value. However, despite the fact that authenticity often might *also* promote happiness, I argue that we seem to value authenticity also in situations where it goes against our happiness. Nozick's (1974) classical thought experiment about an experience machine that would bring as much pleasure as the person wants, is one way to demonstrate this. Nozick (1974, 43) argues that not only pleasure matters in our decision to plug in or not, we also "want to *do* certain things, and not just have the experience of doing them." Furthermore, "we want to *be* a certain way, to be a certain sort of person", instead of "an indeterminate blob." What he thus seems to argue is that besides pleasurable experiences, we also value having authentic experiences and being authentically the person we are. In other words, we seem to value people's right to express themselves and be authentic, even when this leads to people making choices that diminish their happiness.

What about the existence of a basic motivational disposition that would align itself with this self-justifying value for authenticity? Fortunately, there is such a disposition: Autonomy has been defined as being about a sense of volition and a perception of an internal locus of causality (de Charms, 1968; Deci & Ryan, 2000). Autonomy thus means that "one's behaviors are self-endorsed, or congruent with one's authentic interests and values" (Ryan & Deci, 2017, 10). Research within self-determination theory has argued that it is one of the basic psychological needs of human beings, psychological needs being "nutrients that are essential for growth, integrity, and well-being" (Ryan & Deci, 2017, 10). Empirical research has demonstrated the importance of such autonomy for human wellness and vitality (see Martela & Ryan, 2016b; Ryan & Deci, 2017). Research has also shown that autonomy is not only important in Western countries but across the world, also in more collectivistic cultures (e.g. Chen et al., 2015; Chirkov, Ryan, Kim, & Kaplan, 2003)[6]. Indeed, research on human values has shown that when the materialistic conditions improve, cultures across the world tend to move from more survival-related values towards values that put more emphasis on self-expression (Inglehart, Foa, Peterson, & Welzel, 2008; Welzel, 2013). Furthermore, experimental research has shown that when the sense of authenticity is strengthened, this increases people's meaning in life (e.g. Schlegel, Hicks, King, & Arndt, 2011).

In evolutionary terms, the need for autonomy is connected to the propensities in animate life toward "self-regulation of action and coherence in the organism's behavioral aims" (Deci & Ryan, 2000, 253). It is beneficial for the organism to be sensitive to and avoid coercive contexts, as such contexts provide the organism less opportunities to ensure that the situation supports its survival and thriving. As Deci and Ryan (2000, 253) put it "the evolved capacity for autonomy is the means by which humans can avoid having their behavior easily entrained down maladaptive, even disastrous, paths." When autonomous, the organism is better able to regulate its actions in accord with its full array of needs and available capacities, thus leading to more effective self-maintenance. Autonomy as a biological need is thus connected with general evolutionary theory about the organismic tendency to seek increasingly autonomous arrangements (e.g., Rosslenbroich, 2009; Ruiz-Mirazo & Moreno, 2012). However, it should be acknowledged that beyond research within self-determination theory, more research would be needed that would directly connect the human need for autonomy with evolutionary fitness.

[6] It is worth noting that autonomy is not the same thing as individualism, even though they sometimes are confused. As Chirkov et al. (2003, 98) note, autonomy is actually "largely orthogonal to both independence and individualism." While autonomy is about behavior being "willingly enacted" and personally endorsed, individualism is about separateness from others and priority given to personal preferences and goals over collective norms and goals (p. 98, 100). This means that within a collective culture, a person can willingly enact collective goals thus being high in autonomy.

Open questions: Biological and cultural evolution as regards self-justifying values

The present account has concentrated on those motivational dispositions that human beings have acquired through evolution, aiming to identify them, and the corresponding candidates for self-justifying values. The reason to concentrate on them is because such human biological traits would have a good claim to be intuitively appealing in an empirically universal way. However, what such account ignores is the significant influence of cultural evolution on human motivations and values. All conduct is, after all, "*interaction* between elements of human nature and the environment, natural and social" (Dewey, 1922, 10).

First, given that various human societies face relatively similar questions as regards allocation of resources and other often repeated social dilemmas, it could be possible that certain values, although not backed by biological evolution, would be empirically universal just because almost all human societies would have found them to be important in coordinating their behavior. I don't immediately know what could be strong candidates for such culturally but not biologically universal values, but we must remain open for such a possibility.

Second, what is a more probable scenario, a particular culture or a group of cultures might benefit from having a certain self-justifying value, given the state of that society and the internal and environmental pressures they are facing. Thus a particular culture might have to supplement the biologically given self-justifying values with certain more particular self-justifying values. Something akin to the self-justifying value of sacrificing oneself for one's country or the honor bestowed upon those who fight for one's country could perhaps be such a particular self-justifying value that would be beneficial for certain countries in certain historical settings. However, my aim here is not to develop this account nor defend it, it serves just as an example as a possible culturally specific self-justifying value. The more general point is that we must remain open to the (quite probable) possibility that a society might need to supplement the biologically acquired motivational dispositions with culturally enacted values in order to function and thrive. And if a certain value becomes an inherent part of a certain culture and the children are brought to value it from the very beginning, people within that culture might come to value it just as strongly and intuitively as they value the biologically based motivational dispositions. Thus, within cultures there could be culturally acquired self-justifying values that could be widely shared and have just as strong intuitive appeal and motivational force as the biological dispositions, as long as we stay within the boundaries of that particular culture.

Third, we must also remain open to the possibility that in a particular culture, the biologically acquired motivational dispositions and the culturally acquired values could be in open conflict. The culture might see a need to value a thing the valuing of which diminishes the possibility to value some of the self-justifying values identified above. For example, a culture might see it important to value homogeneity to such a degree that it narrows significantly people's room for authenticity and self-expression. How to reconcile such conflicts between biologically and culturally acquired values? A full exposition of this question would require an article of its own, but suffice it to say here that while there might be cases where the culturally acquired value might be completely legitimate, given the specific environmental and other factors, often it is useful to think whether the value that is in conflict with our biological dispositions actually serves the good of the whole society. For example, the narrow space for self-expression that many cultures allow for women usually doesn't serve the good of all the people of the culture but rather the interests of those in power. Thus these conflicts must be resolved case by case to see how strong case can be made for each of the value in question in that particular historical setting. But in general, given their empirical universality, the biologically acquired dispositions and their corresponding self-justifying values have the potential to be used to evaluate and compare cultures and cultural practices 'from the outside.' Thus they could offer an important tool that could be used to evaluate and criticize certain cultural practices and values within certain tribes, organizations, or societies (see Ryan & Deci, 2017).

All in all, what I am trying to say is that although the empirical identification of basic motivational dispositions and corresponding self-justifying values provides one very promising avenue for identifying those self-justifying values that would have very wide appeal for human beings, they alone are not enough to settle the most suitable values for individuals or societies. In any particular life or in any particular society valuing them must be balanced with an examination of what more particular values or motives might be important for survival and thriving in that specific historical setting.

Conclusion

This article has aimed to articulate one path through which human quest to find warranted and intuitively appealing self-justifying values to guide the life of individuals and societies could be brought closer to an empirical science. Instead of just appealing to intuition, the present account suggests that one robust and empirically universal source of such intuitions are the basic motivational dispositions humans are equipped with through their biological nature. Psychological and evolutionary research has examined the potential candidates for such dispositions, providing us with a relatively broad body of empirical research that can be used to evaluate the strength of the case behind any candidate motivational disposition.

Some eighty years ago, Dewey (1939, 21) noted how medical art is approaching

> a state in which many of the rules laid down for a patient by a physician as to what it is better for him to do, not merely in the way of medicaments but of diet and habits of life, are based upon experimentally ascertained principles of chemistry and physics.

Dewey called for a similar turn in the science of human valuations. When we philosophers, as experts of ethical matters, are asked to give advice for individuals or for societies about what they ought to value, what can we base those advices on? The possibility suggested here is that empirical examination of human basic motivational dispositions could offer us one tool to be used for grounding such advice in something empirically universal. Although the particularities of individual lives and societies must be taken into account, the self-justifying values identified through the empirical method suggested here would be values that it is wise to acknowledge in almost all cases. One might have to balance them with some more particular values, but there would be almost no cases where these self-justifying values would not play any role in identifying what it is best for a person or a society to do. Thus they might represent the most robust advice that one can give to the general question of what to value and what is worth striving for in human life.

Aalto University

References

Aknin, L. B., Barrington-Leigh, C. P., Dunn, E. W., Helliwell, J. F., Burns, J., Biswas-Diener, R., … Norton, M. I. (2013), "Prosocial spending and well-being: Cross-cultural evidence for a psychological universal". *Journal of Personality and Social Psychology*, *104*(4), pp. 635–652.

Aristotle. (2012), *Nicomachean ethics*. (R. C. Bartlett & S. D. Collins, Trans.). Chicago: University of Chicago Press.

Bailey, D. H., & Geary, D. C. (2009), "Hominid Brain Evolution: Testing Climatic, Ecological, and Social Competition Models". *Human Nature : An Interdisciplinary Biosocial Perspective; New York*, *20*(1), pp. 67–79. https://doi.org/http://dx.doi.org/10.1007/s12110-008-9054-0

Batson, C. D., Ahmad, N., Yin, J., Bedell, S. J., Johnson, J. W., & Templin, C. M. (1999), "Two Threats to the Common Good: Self-Interested Egoism and Empathy-Induced Altruism". *Personality and Social Psychology Bulletin*, *25*(1), pp. 3–16. https://doi.org/10.1177/0146167299025001001

Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991), "Empathic joy and the empathy-altruism hypothesis". *Journal of Personality and Social Psychology*, *61*(3), pp. 413–426.

Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995), "Immorality from empathy-induced altruism: When compassion and justice conflict". *Journal of Personality and Social Psychology*, *68*(6), pp. 1042–1054.

Baumeister, R. F., & Leary, M. R. (1995), "The need to belong: Desire for interpersonal attachments as a fundamental human motivation". *Psychological Bulletin*, *117*(3), pp. 497–529.

Becker, L. C. (1992), "Good lives: prolegomena". *Social Philosophy and Policy*, *9*(2), pp. 15–37.

Bentham, J. (1789), *An Introduction to the Principles of Morals and Legislation*. London: T. Payne & Son.

Boyd, R., Gintis, H., & Bowles, S. (2010), "Coordinated punishment of defectors sustains cooperation and can proliferate when rare". *Science*, *328*(5978), pp. 617–620.

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Deeder, J., … Verstuyf, J. (2015), "Basic psychological need satisfaction, need frustration, and need strength across four cultures". *Motivation and Emotion*, *39*(2), pp. 216–236.

Chirkov, V., Ryan, R. M., Kim, Y., & Kaplan, U. (2003), "Differentiating autonomy from individualism and independence: A self-determination theory perspective on internalization of cultural orientations and well-being". *Journal of Personality and Social Psychology*, *84*(1), pp. 97–110.

Cottingham, J. (2003), *On the Meaning of Life*. London: Routledge.

Crumpler, H., & Grossman, P. J. (2008), "An experimental test of warm glow giving". *Journal of Public Economics*, *92*(5–6), pp. 1011–1021. https://doi.org/10.1016/j.jpubeco.2007.12.014

Damon, W. (2008), *The path to purpose: How young people find their calling in life*. New York: Free Press.

de Charms, R. (1968), *Personal causation*. New York: Academic Press.

De Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004), "The neural basis of altruistic punishment". *Science*, *305*(5688), pp. 1254–1258.

de Waal, F. B. M. (2009), *Primates and Philosophers: How Morality Evolved*. Princeton University Press.

Decety, J., & Cowell, J. M. (2015), "Empathy, justice, and moral behavior". *AJOB Neuroscience*, *6*(3), pp. 3–14.

Deci, E. L., & Ryan, R. M. (2000), "The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior". *Psychological Inquiry*, *11*(4), pp. 227–268.

Dewey, J. (1922), *Human nature and conduct*. New York: Henry Holt and Company.

Dewey, J. (1929), *The Quest for Certainty*. New York: Minton, Balch & Co.

Dewey, J. (1932), *Ethics - Revised Edition*. New York: Henry Holt and Company.

Dewey, J. (1938), *Logic - The Theory of Inquiry*. New York: Henry Holt and Company.

Dewey, J. (1939), *Theory of valuation*. Chicago, Ill.: University of Chicago Press.

Dunbar, R. I. M. (1998), "The social brain hypothesis". *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), pp. 178–190.

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008), "Spending money on others promotes happiness". *Science*, *319*(5870), pp. 1687–1688.

Engel, C. (2011), "Dictator games: a meta study". *Experimental Economics*, *14*(4), pp. 583–610.

Fehr, E., & Fischbacher, U. (2003), "The nature of human altruism". *Nature*, *425*, pp. 785–791.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005), "Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence". *Evolution and Human Behavior*, *26*(1), pp. 10–46.

Gintis, H., Smith, E. A., & Bowles, S. (2001), "Costly signaling and cooperation". *Journal of Theoretical Biology*, *213*(1), pp. 103–119.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013), "Moral foundations theory: The pragmatic validity of moral pluralism". In *Advances in Experimental Social Psychology,* San Diego, Ca.: Academic Press, pp. 55–130.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011), "Mapping the moral domain". *Journal of Personality and Social Psychology*, *101*(2), pp. 366–385.

Haidt, J. (2001), "The emotional dog and its rational tail: A social intuitionist approach to moral judgment". *Psychological Review, 108,* pp. 814–834.

Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007), "Neural responses to taxation and voluntary giving reveal motives for charitable donations". *Science, 316*(5831), pp. 1622–1625.

Hardy, C. L., & Van Vugt, M. (2006), "Nice guys finish first: The competitive altruism hypothesis". *Personality and Social Psychology Bulletin*, *32*(10), pp. 1402–1413.

Haybron, D. M. (2000), "Two philosophical problems in the study of happiness". *Journal of Happiness Studies*, *1*(2), pp. 207–225.

Haybron, D. M. (2008), *The pursuit of unhappiness: the elusive psychology of well-being*. New York: Oxford University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001), "In search of homo economicus: behavioral experiments in 15 small-scale societies". *American Economic Review*, *91*(2), pp. 73–78.

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

Inglehart, R., Foa, R., Peterson, C., & Welzel, C. (2008), "Development, freedom, and rising happiness: A global perspective (1981–2007)". *Perspectives on Psychological Science*, *3*(4), pp. 264–285.

Janoff-Bulman, R., & Carnes, N. C. (2013), "Surveying the Moral Landscape: Moral Motives and Group-Based Moralities". *Personality and Social Psychology Review*, *17*(3), pp. 219–236.

Jensen-Campbell, L. A., Graziano, W. G., & West, S. G. (1995), "Dominance, prosocial orientation, and female preferences: Do nice guys really finish last?" *Journal of Personality and Social Psychology*, *68*(3), pp. 427–440.

Kauppinen, A. (2016), "Meaningfulness". In G. Fletcher (Ed.), *The Routledge Handbook of Philosophy of Well-Being* (pp. 281–291). Abingdon, UK: Routledge.

Kierkegaard, S. (1992), *Concluding Unscientific Postscript to "Philosophical Fragments", Vol. 1*. Princeton, NJ.: Princeton University Press.

Luccasen, A., & Grossman, P. J. (2017), "Warm-glow giving: Earned money and the option to take". *Economic Inquiry*, *55*(2), pp. 996–1006.

Mandela, N. (1995), *Long Walk to Freedom: The Autobiography of Nelson Mandela*. Boston: Little, Brown and Company.

Marsh, A. A. (2016), "Neural, cognitive, and evolutionary foundations of human altruism". *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(1), pp. 59–71.

Martela, F. (2015), "Fallible inquiry with ethical ends-in-view: A pragmatist philosophy of science for organizational research". *Organization Studies*, *36*(4), pp. 537–563.

Martela, F. (2017a), "Meaningfulness as Contribution". *The Southern Journal of Philosophy*, *55*(2), pp. 232–256. https://doi.org/10.1111/sjp.12217

Martela, F. (2017b), "Moral Philosophers as Ethical Engineers: Limits of Moral Philosophy and a Pragmatist Alternative". *Metaphilosophy*, *48*(1–2), pp. 58–78. https://doi.org/10.1111/meta.12229

Martela, F. (2018), "Intrinsic values grounded in basic motivational dispositions – How to be a subjectivist about meaning in life". *Under Review*.

Martela, F., & Ryan, R. M. (2016a), "Prosocial behavior increases well-being and vitality even without contact with the beneficiary: Causal and behavioral evidence". *Motivation and Emotion*, *40*(3), pp. 351–357.

Martela, F., & Ryan, R. M. (2016b), "The Benefits of Benevolence: Basic Psychological Needs, Beneficence, and the Enhancement of Well-

Being". *Journal of Personality*, *84*(6), pp. 750–764. https://doi.org/10.1111/jopy.12215

Maslow, A. H. (1968), *Toward a psychology of being (2nd ed.)*. New York: Van Nostrand.

May, T. (2015), *A significant life - Human meaning in a silent universe*. Chicago: University of Chicago Press.

Metz, T. (2012), "The meaningful and the worthwhile: Clarifying the relationships". *The Philosophical Forum*, *43*(4), pp. 435–448.

Metz, T. (2013), *Meaning in life: An Analytic Study*. Oxford: Oxford University Press.

Mill, J. S. (1863), *Utilitarianism*. London: Parker, Son and Bourn.

Moore, G. E. (1903), *Principia ethica*. Cambridge: Cambridge University Press.

Niemiec, C. P., Ryan, R. M., & Deci, E. L. (2009), "The path taken: Consequences of attaining intrinsic and extrinsic aspirations in post-college life". *Journal of Research in Personality*, *43*(3), pp. 291–306.

Nietzsche, F. W. (1961), *Thus spoke Zarathustra*. (R. J. Hollingdale, Trans.). London: Penguin Books.

Nowak, M. A., & Sigmund, K. (2005), "Evolution of indirect reciprocity". *Nature*, *437*(7063), pp. 1291–1298.

Nozick, R. (1974), *Anarchy, state, and utopia*. Padstow: Blackwell.

Prinz, J. (2008), *The Emotional Construction of Morals*. New York: Oxford University Press.

Rogers, C. (1961), *On becoming a person: A therapist's view of psychotherapy*. Boston: Houghton Mifflin.

Rosslenbroich, B. (2009), "The theory of increasing autonomy in evolution: a proposal for understanding macroevolutionary innovations". *Biology & Philosophy*, *24*(5), pp. 623–644.

Ruiz-Mirazo, K., & Moreno, A. (2012), "Autonomy in evolution: from minimal to complex life". *Synthese*, *185*(1), 21–52.

Ryan, R. M., & Deci, E. L. (2017), *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York: Guilford Press.

Ryan, R. M., & Hawley, P. H. (2017), "Naturally good? Basic psychological needs and the proximal and evolutionary bases of human benevolence". In K. W. Brown & M. R. Leary (Eds.), *The Oxford handbook of hypo-egoic phenomena*, New York: Oxford University Press, pp. 205–222.

Sartre, J.-P. (2007), "Existentialism is a humanism". In C. Macomber (Ed. & Trans.), *Existentialism is a Humanism* (pp. 17–72). New Haven, CT: Yale University Press.

Schlegel, R. J., Hicks, J. A., King, L. A., & Arndt, J. (2011), "Feeling like you know who you are: Perceived true self-knowledge and meaning in life". *Personality and Social Psychology Bulletin*, *37*(6), pp. 745–756.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997), "The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering". In A. Brandt & P. Rozin (Eds.), *Morality and health,* New York: Routledge, pp. 119–169.

Smuts, A. (2013), "The Good Cause Account of the Meaning of Life". *The Southern Journal of Philosophy*, *51*(4), pp. 536–562.

Street, S. (2006), "A Darwinian dilemma for realist theories of value". *Philosophical Studies*, *127*(1), pp. 109–166.

Taylor, C. (1991), *The ethics of authenticity*. Cambridge, Mass.: Harvard University Press.

Tolstoy, L. (2000), "My Confession". In E. D. Klemke (Ed.), L. Wierner (Trans.), *The Meaning of Life - Second Edition*, New York: Oxford University Press, pp. 11–20.

Warneken, F., & Tomasello, M. (2015), "The developmental and evolutionary origins of human helping and sharing". In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford Handbook of Prosocial Behavior*, Oxford: Oxford University Press, pp. 100–113.

Wedekind, C., & Milinski, M. (2000), "Cooperation through image scoring in humans". *Science*, *288*(5467), pp. 850–852.

Welzel, C. (2013), *Freedom rising - Human empowerment and the quest for emancipation*. New York: Cambridge University Press.

West, S. A., El Mouden, C., & Gardner, A. (2011), "Sixteen common misconceptions about the evolution of cooperation in humans". *Evolution and Human Behavior*, *32*(4), pp. 231–262.

Wolf, S. (1997), "Happiness and meaning: two aspects of the good life". *Social Philosophy and Policy*, *14*(1), pp. 207–225.

Wolf, S. (2016), "Meaningfulness: A Third Dimension of the Good Life".
Foundations of Science, *21*(2), pp. 253–269.
https://doi.org/10.1007/s10699-014-9384-9

What are Tropes, Fundamentally?
A Formal Ontological Account[1]

JANI HAKKARAINEN

Introduction

Consider the rest mass of an electron and the equal rest mass of another electron. According to a metaphysical theory called "trope theory", the rest mass of each electron is an entity numerically distinct from both the two electrons and the other rest mass. The rest masses are not numerically identical; they are tropes. Accordingly, tropes are routinely called "particular properties" or "particularized properties" in the metaphysical literature (e.g. Garcia 2016, 2). Introductions to metaphysics in particular discuss tropes as such entities.[2] This is intimately connected to the dichotomous approach to tropes whereby they must be seen as either properties or objects (i.e. bearers of properties), or as something akin to one or other of these options (Maurin 2018, sec. 2.1). For example, David M. Armstrong famously calls tropes "junior substances"; hence they are more akin to objects than properties (1989, 115).

In this paper, I argue that when one considers the basics of trope theory, one is on the wrong track right from the start when using this dichotomous set-up.[3] The set-up is deeply misleading when one tries to understand what tropes are

[1] I would like to thank the audience at PSFC 2018, the anonymous referees of the *Acta* and Markku Keinänen for comments.

[2] Cf. Allen 2016, 39–40; Edwards 2014, 49; Effingham, Beebee & Goff 2010, 255.

[3] By "trope theory", I refer to the trope bundle theories of substances and objects, in contrast to their trope substratum theories. "Trope nominalism" covers both of these.

fundamentally. Here I use "fundamentally" in formal ontological terms; this means, to a first approximation, the fundamental *form of existence* of tropes. I shall present the funfundamental form of existence of tropes as it is represented by the Strong Nuclear Theory (SNT) of tropes and substances developed by Markku Keinänen and the present author.[4] The SNT states that the full fundamental form of existence of each trope – that is, its full fundamental *ontological form* – is to be a strongly rigidly or generically dependent (mereologically) simple individual part. Neither propertyhood nor objecthood is mentioned here. The same result should concern any trope theory as a bundle construction of objects.

The SNT distinguishes the ontological form of a trope from the identification of the trope with a nature or character (Hakkarainen & Keinänen 2017). Ontologically, each trope is an entity identified with a nature or character. Since the SNT involves the distinction between ontological form and ontology, it is construed in a specific metaphysical tradition I call the "formal ontological". The formal ontological tradition stems from Edmund Husserl's *Logical Investigations* (1900–1), but it was initiated in analytic metaphysics by Barry Smith and Kevin Mulligan (Smith 1978; 1981; Smith & Mulligan 1983). The basic idea of the formal ontological tradition is that the primary subject matter of metaphysics is ontological form, which includes the membership of ontological categories. Formal ontology studies both. Ontological form provides a unique point of view to the other main branch of metaphysics, namely ontology, which studies questions of existence, such as whether there are abstract entities or properties – that is, members of certain putative categories.

Nonetheless, no fully satisfying account of ontological form and its difference from existence or being has so far been put forward in the formal ontological literature. Therefore, there is a dire need for a *metatheory of formal ontology* in which this deficiency is resolved. Furthermore, a fully satisfactory account of ontological form has to include a sophisticated view of fundamentality and non-fundamentality, which

[4] Keinänen 2011; Keinänen & Hakkarainen 2010; 2014; Hakkarainen & Keinänen 2017; cf. also Keinänen, Keskinen & Hakkarainen 2017.

are intensively discussed by metaphysicians and metametaphysicians (as is documented by Tahko 2018).

Accordingly, the SNT as a formal ontology needs to be elaborated upon by my metatheory. This elaboration, which is the aim of my paper, especially concerns fundamental ontological form. Therefore, I mostly assume the SNT and do not defend its central tenets here.

The paper has a six-part structure. To describe the fundamental ontological form of tropes in the SNT, first I have to go into a rather long discussion of my metatheory. This I do in the first two sections of the paper. In the third section, I apply my metatheory to the SNT, which leads me to argue in the fourth section that the dichotomous set-up of properties or objects (or something akin to one or the other) is a nonstarter in the SNT when one considers the fundamental ontological form of tropes. With the help of my metatheory, in the fourth section I also establish that the arguments against tropes by Herbert Hochberg (2004), Douglas Ehring (2011) and Robert K. Garcia (2014b; 2015; 2016) fail. This section thus shows the fruitfulness of the elaboration of the SNT by my metatheory. The fifth section discusses two non-fundamental ontological forms of tropes in the SNT: proper parthood of substances and concreteness. I wrap things up in the sixth section with the conclusion.

1. Ontological Form as Distinguished from Being or Existence

To distinguish ontology from formal ontology in a determinate manner, I have to make a clear and precise distinction between being and ontological form. Regarding being, it is not necessary to go into the numerous questions concerning it, such as whether it is to be expressed by a quantifier or predicate. Suffice it to say, I simply make two assumptions about being in this paper, leaving room for more than one view of it. (1) "Being" and "existence" are both univocal. (2) I follow the mainstream view in analytic metaphysics and metaontology that "existence" and its cognates are interchangeable with "being" and its cognates (van Inwagen 2009).

Let me introduce, for the theoretical purposes below, the technical primitive concept of *character* at this point: the character of an entity is *what the entity is like*. Paradigmatic examples of characters are the qualities and quantities entities presumably are or have, such as shape and rest mass. So, *character* covers tropes, accidents, attributes and properties, and it is therefore a more general concept than all of these.[5] In principle, a character may be essential, necessary or contingent to an entity. Therefore, the concept of character here also differs from the concept of essence, regardless of whether essence is understood modally or non-modally. The characters of entities belong to the extension of the concept of being or existence in the metaphysical theories that are committed to the existence of characters – for example, realism about property universals, mereological nominalism, class or set nominalism and trope nominalism. The upshot is that being consists of entities, including their character, given there are any characters or entities have any character.

The concept of *ontological form*, in turn, is a complex concept consisting of the concepts of *being* or *existence* and *form.* I have a relational account of form in terms of the concept of *character-neutral internal relation*. Character-neutral relations are internal because they are not entities numerically distinct from their relata ("additional entities" in this sense). The relational terms occurring in statements about internal relations do not designate (name) any relational entity (Keinänen, Keskinen & Hakkarainen 2017, ch. 2). They only *apply* to the relata of the internal relation: their reference is divided. To say that books are numerically distinct is not to name any entity additional to the books. Rather, it is to apply numerical distinctness to the books. Yet the holding of the internal relations of their relata may be in principle asserted by relational statements expressing propositions true of the relata, such as "the books are numerically distinct" (ibid.). So, in this specific sense, the holding of internal relations is real: the books, for instance, really are numerically distinct.

Character-neutrality is independence from what an entity is like. Thus, a character-neutral internal relation holds inde-

[5] As a consequence, being a bare particular/substance or haecceity is not a character. Rather, it is an ontological form.

pendent from the character of its relata. When its holding is asserted, the statement as such, even if true, does not say anything whatsoever about the character of the relata. Therefore, character-neutral internal relations are such internal relations whose holding is expressible by true relational statements that do not describe the character of the relata without further assumptions. Hence, I may initially say that the ontological form of entities is determined by their standing in character-neutral internal relations.

To argue this, let us consider four examples that are typically discussed by contemporary metaphysicians: *being numerically distinct from*, *depending ontologically on*, *being a whole of* and *being a proper part of*. Each of these is *relational*: they are features that entities have in virtue of being related to something; for instance, *x* is a whole in virtue of being related to some entities – that is, to its proper parts. These relational features of entities may also be tentatively characterized as *ways in which entities exist*: *x* exists *as* numerically distinct from *y*, *x* exists as ontologically dependent on *y*, *x* exists as a whole of *y* and *z* and *x* exists as a proper part of *y*. Thus, these four features may be said to be the *relational ways of existence* of entities – the existence of entities is their standing in relation to something.

Here we have "way" in the sense of "form"; in these examples we are speaking about the specific *form* of the existence of *x*. Therefore, I may say that the relational way of existence of *x* is its *relational form of existence*. For example, the numerical distinctness of a book from other books is its relational way of existence, rather than its character.

The four relational forms of existence above are character-neutral, which can be seen by considering the statements that *x* is numerical distinct from *y*, that *x* ontologically depends upon *y*, that *x* is a whole of *y* and *z*, and that *x* is a proper part of *y*. None of them, *without further assumptions*, describe the character of *x*, *y* or *z* at all. As such, it does not tell us anything about the character of *x*, *y* and *z* that *x* is numerically distinct from *y*, that *x* ontologically depends upon *y*, that *x* is a whole of *y* and *z* or that *x* is a proper part of *y*. Therefore, these are *formal* statements: they concern the relational form of existence of their relata.

The etymological origin of "ontological" is the Greek *ontos*, which can be translated into the possessive form of "existence". So, a general concept that covers these four typical examples is *ontological form*: standing in certain character-neutral relations. *Being numerically distinct from*, *depending ontologically on*, *being a whole of* and *being a proper part of* are ontological forms. Other plausible candidates for typical examples of ontological forms in different metaphysical theories are *being numerically identical to, being a part of, being a member of, being an element of, instantiating, exemplifying, participating, modifying, characterizing* and different *types* of *depending ontologically on*, such as *depending for its existence rigidly or non-rigidly on* (cf. Tahko & Lowe 2015).

Let us follow the clue of these four typical, paradigmatic examples. Since they are paradigmatic, they generalize: true relational statements about ontological forms do not say anything about the character of entities without further assumptions. Ontological forms of entities consist of or may be construed as their standing in character-neutral relations. Since the order of character-neutral relations might make a difference, the order is to be considered. Proper parthood, for instance, is asymmetric (and standardly dyadic). Furthermore, it is E.J. Lowe's insight that ontological forms are better considered internal rather than external relations on pain of a vicious infinite regress (2006, 80, 92, 111, 167). Therefore, I can conclude that *for an entity to have an ontological form is for it to be a relatum of a character-neutral internal relation or relations jointly in an order*.[6]

The "is" in the previous statement is neither predication nor numerical identity. It is "the is of generic identity": (for an entity) to have an ontological form *is generically identical with* (for it) to be a relatum of a character-neutral internal relation or relations jointly in an order. So, I need to introduce the notion of generic identity next. This notion will turn out

[6] As such, ontological form differs from the possible logical form; to a first approximation, ontological form concerns entities, whereas logical form concerns truths or truth-bearers *qua* true or false (cf. Smith & Mulligan 1983, 73). Thus, logical connectives such as negation and disjunction are not formal ontological, although they might be character-neutral. It is a different metaphysical question as to whether there are corresponding formal ontological concepts.

to be crucial also for my account of formal ontological fundamentality and non-fundamentality, including the fundamental and non-fundamental ontological forms of tropes below. Therefore, generic identity needs to be elucidated, although I assume it is a primitive notion.

Generic identity is a form of *generalized identity*, which is a newcomer notion in philosophy, although its plausible examples are familiar: for instance, "for an entity to be a bachelor is for it to be an unmarried adult male" and "for an entity to be a water molecule is for it to be an H_2O molecule". Groundbreaking work on generalized identity has been done by Augustin Rayo (2013), Øystein Linnebo (2014), who coined the term, Cian Dorr (2016), Fabrice Correia (2017), and Correia and Alexander Skiles (2017).

I follow Correia and Skiles and consider generalized identity analogous to familiar numerical or *objectual identity* (e.g. "Hesperus is Phosphorous"). Correia and Skiles (2017, 3) express generalized identity with an operator, ≡, indexed by one or more variables, which takes two open or closed sentences. *Generic identity* is generalized identity of the form "for an entity to be F is for it to be G" in the monadic case ($Fx \equiv_x Gx$), which can be generalized into polyadic cases that involve relational predicates such as character-neutral internally relational terms. Generic identity, just like objectual identity, is reflexive, symmetric and transitive (Correia & Skiles 2017, 4, 8). It has transparent linguistic contexts concerning only metaphysical matters rather than their mode of presentation (Dorr 2016, 44; Correia & Skiles 2017, 4).

The expressions flanking ≡ can be conjunctive (Correia & Skiles 2017, 2). Still, Correia and Skiles (2017, 3) emphasize that a generic-identity statement as such does not commit us to the existence of conjunctive properties or facts, which some might find metaphysically problematic. Unlike objectual identity, the terms of generic identity do not have to be entities or its sign's flanking expressions designating true or satisfied (ibid.). For example, it may hold of "for an entity to be a bachelor" and "for it to be an unmarried adult male" even if there were no bachelors, that is, unmarried adult males.

Generic identity allows for *representational* differences between the left-hand side and the right-hand side of ≡, as well as in the objectual identity "Virginia Woolf is Virginia

Stephen" (arguably these names differ in meaning). So representational asymmetry is possible and the right-hand side may be informative about the left-hand side. Thus, the generic identity of the ontological form of an entity with its standing in a character-neutral internal relation or relations jointly in an order may very well be symmetric *and* informative.

Since for an entity to have an ontological form is for it to be a relatum of a character-neutral type of internal relation or relations jointly in an order, the suitable term for these ontological forms is "formal ontological relation" (FOR; cf. Smith & Grenon 2004, Lowe 2006, ch. 3).[7] Accordingly, true formal ontological relational statements do not tell us anything about the character of their relata without further assumptions. Rather, they describe the character-neutral relational way in which the relata exist. Hence, for an entity to have an ontological form is for it to be a relatum of a FOR or FORs jointly in an order. In Aristotelian realism, for instance, for an entity to have the ontological form of being a universal is for it to be a terminus of the FOR of instantiation.

By contrast, neither indistinguishability, exact resemblance/similarity, (inexact) resemblance/similarity nor any of their opposites is a FOR. They are character-*dependent* internal relations. Their statements even without further assumptions tell us something about the character of their relata. Let us assume it is true that *x* exactly resembles *y* and we know it. This true statement as such says something about the character of *x* and *y*, namely, that they are exactly resembling; the statement could not be true without something being true of the character of *x* and *y*.

On this basis, I am also able to draw a clear-cut distinction between formal ontological and other internally relational *terms*. Formal ontological terms are character-neutral internally relational terms, whereas other internally relational terms are character-dependent: they appear in statements that in themselves say at least something about the character of the entities to which they apply. Moreover, formal ontological terms are *primitive* if they cannot be non-circularly de-

[7] Therefore, generic identity is not a FOR, since the terms of generic identity do not have to be entities, in contrast to internal relations such as FORs.

fined. *Derivative* formal ontological terms, in turn, may be non-circularly defined. It depends on the metaphysical theory as to which formal ontological terms are primitive and which derivative. For instance, "is a part of" is considered primitive and "is a proper part of" derivative (and dyadic) in the metaphysical theories following the standard axiomatization of classical mereology.

2. Formal Ontological Relations and Formal Ontological Fundamentality

2.1 The Types and Ground of Formal Ontological Relations

If FORs are character-neutral internal relations, why do they hold of their relata if they do? Why do entities have the ontological forms they have? To answer this question, I first need to elaborate on the sense in which FORs are internal. This involves drawing important distinctions between different types of FORs.

Due to their character-neutrality, FORs cannot be internal in the "property conception of internal relations" sense – as held by Armstrong (1989, 43), for example – which grounds the holding of any internal relation in the character of its relata. FORs are internal by the "modified existential conception" of internal relations, for which I have argued elsewhere (Hakkarainen, Keinänen & Keskinen 2018, 93–102; cf. Keinänen, Keskinen & Hakkarainen 2017, ch. 2).

This modified existential conception elaborates upon Mulligan's existential account, according to which the mere joint existence of the relata of an internal relation is sufficient and exhaustively necessary for its holding (1998, 344). Mulligan's existential conception needs to be modified to cover a plausible key *type* of internal relations in metaphysical literature. Here, we are speaking about the situation where the mere existence of the relata is jointly sufficient for the holding of a relation, but the existence of entities distinct from the relata of the relation is also necessary for its holding. Indeed, it is better to consider this kind of case an internal rather than an external relation in order to avoid worries about Bradley's relation regress threatening the latter but not the former (Lowe 2006, 111; Hakkarainen & Keinänen 2016).

This type of case may be illustrated by qualitative and quantitative relations among objects in views that are committed to properties. Let us take the exact resemblance of objects as an example. Assume for the sake of argument that electrons are objects and the electron charge (-e) is *their* essential or *de re* necessary (particular or universal) property numerically distinct from the electrons. Independent of the details of the metaphysical description of this circumstance, the sole existence of two electrons is jointly sufficient and individually necessary for the holding of the internal relation of having the same charge as between the two electrons. Necessarily, if these electrons exist, then they have the property of -e charge and the same charge. Due to this sufficiency, there is no ontological need to reify the relation of having the same charge as into an external relation. However, the existence of entities numerically distinct from the relata – that is, the two electrons – is also necessary for the holding of the relation. The existence of the property of -e is necessary for the holding of the relation of having the same charge as. The necessity basis for this holding includes the property in addition to the two electrons (other examples are provided by proper parthood (given certain assumptions) and Lowe's FOR of exemplification below).

To cover these important cases, the modified existential conception makes a tripartite distinction among internal relations. When this distinction is elaborated on by the concept of generic identity and applied to FORs for the present purposes, it reads as follows. In the first place, for the holding of *proto* FORs, the mere existence of their relata is jointly sufficient and individually necessary. Secondly, a distinction between *derived* FORs and *basic* FORs is partly put in terms of proto FORs:

> [DFOR]: Necessarily, entities a_1, …, a_n stand in *derived FOR* R if and only if the holding of R of a_1, …, a_n is generically identical to the joint holding of proto FORs holding between entities some of which are distinct from a_1, …, a_n [a_1, …, a_n are names of entities]

> [BFOR]: Necessarily, entities a_1, …, a_n stand in *basic FOR* R if and only if R is a proto FOR and the holding of R of a_1, …, a_n is not

generically identical to the joint holding of proto FORs holding between entities some of which are distinct from $a_1, \ldots, a_n$.

The basic and derived FORs are mutually exclusive and jointly exhaustive. All *basic* FORs are proto internal – for instance primitive (inexplicable) numerical identity for the holding of which the mere existence of primitively numerically identical entities is sufficient and exhaustively necessary. A putative theoretical example of a derived FOR is Lowe's exemplification between a substance (e.g. Dobbin the horse) and a universal property (e.g. warm-bloodedness or whiteness) (2006, 40, 92–3, 95, 206). In the four-category ontology, its holding may be construed as being generically identical to the joint holding of either

(1) instantiation between the universal property (e.g. whiteness) and a mode (being white) *and* characterization between the mode and the substance (e.g. Dobbin), or

(2) instantiation between the substance and a kind (horse) *and* characterization between the kind and the universal property (e.g. warm-bloodedness; ibid.).

If the holding of this exemplification is necessary for the existence of the substance and the universal property, it is a derived proto FOR. In the case that it is only contingent to them, then it is a merely derived FOR.

Thus, among *proto FORs*, there is a further distinction between the basic and the derived. In order for a basic FOR to hold, there need not be any specific entities distinct from the relata (e.g. primitive numerical identity). *Derived proto* FORs hold of their relata in virtue of proto FORs holding between entities some of which are distinct from the relata. The existence of the relata of a derived proto FOR necessitates the existence of entities distinct from the relata. The necessary form of Lowe's exemplification relation is a theoretical example of such a relation. Yet this derived FOR is proto formal ontological because the existence of their relata is jointly sufficient and individually necessary for the holding of these relations.

Consequently, *the necessity and sufficiency basis* for the holding of a FOR depends on the type of the relation. The mere existence of the relata of a basic FOR is both jointly sufficient and exhaustively necessary for its holding (e.g. primitive numerical identity). In the case of a derived proto FOR, the mere existence of its relata is jointly sufficient and individually but not exhaustively necessary for its holding. The existence of entities distinct from the relata is also individually necessary. The holding of the derived proto FOR is generically identical to the joint holding of some proto FORs that bring in additional necessary entities. These additional relata complete the necessity basis of the holding of the derived proto FOR. Again, the necessary form of Lowe's exemplification relation is a derived proto FOR.

If a FOR is merely derived (like the contingent form of Lowe's exemplification), then the existence of its relata is not jointly sufficient for its holding; the sufficiency (and necessity) basis needs to be supplemented by the existence of entities distinct from the relata. The holding of such a derived FOR is contingent upon the existence of its relata. This is made possible by the fact that the holding of the derived FOR is generically identical to the joint holding of some proto FORs that add relata entities. It is the joint existence of the relata of all these proto FORs that is sufficient and exhaustively necessary for the holding of the derived FOR.

To explicate this distinction further, recall that a generic-identity statement does not involve any commitment to the existence of conjunctive properties or facts, which some might find metaphysically problematic. In contrast to objectual identity, the relata of generic identity do not have to be entities or the expressions flanking ≡ designating, satisfied or true. Hence, the holdings of FORs can be generically identical to each other although FORs are not entities that the flanking expressions could designate – especially entities in the category of universal relational properties that are instantiated. Furthermore, the FORs do not have to be particular relational entities of which the flanking expressions are true (e.g. relational tropes). *The generic identity of FORs is the sameness of the really holding character-neutral relatednesses of entities.* Note also that generic identity allows representational differences between the left-hand side and the right-hand side of ≡. Thus, the generic identity of the holding of a derived FOR with the joint holding of proto FORs may very well be symmetric *and* informative of the derived FOR. Furthermore, if generic identity is conjunctive (like in a derived FOR), then the conjuncts can be individually more fundamental in some respect than the other side (cf. Dorr 2016, 43).

By means of the tripartite distinction, I can answer the question with which I began this section: why do FORs hold? Whether a FOR is basic, derived proto or merely derived, its holding boils down to the existence of some entities. Their existence jointly necessitates and is exhaustively necessary for the holding of the FOR. In the case of the basic FORs, these entities are the relata of the FOR. If the FOR is derived, there has to be at least one entity distinct from the relata that is necessary for the holding of the FOR. Depending on whether this additional entity is only necessary or plays a part in completing the sufficiency basis for the holding of the FOR, the FOR is either derived proto or just derived. Be that as it may, *the ground for the holding of a FOR consists only of the existence of entities, rather than their character*. This is how it ought to be given my view that FORs are character-neutral internal relations.

Of these *de re* modalities, one may in principle hold any of the following three alternative metaphysical views. Let me facilitate my expression and focus on the necessity of the holding of a basic FOR upon the existence of its relata. (1) One may defend the view that this necessity is reducible to the existence of the relata of the basic FOR in possible worlds, of which there are several accounts available in the literature (for a mapping of alternatives, cf. Divers 2002). (2) One may take the necessity in question as a primitive fact: it is just an inexplicable brute fact that the existence of the relata is sufficient for the holding of the basic FOR (e.g. primitive numerical identity). (3) One grounds the necessity of the holding of a basic FOR in the inexplicable formal essence of its relata, or at least one of them. We can read Lowe holding this view (2012, 241–3). Although I am leaning towards the second, primitivist view, I do not want to take any firm stance on this issue in the paper. I simply want to point out that my view of the ground of the holding of FORs is available to the upholders of more than one form of modal metaphysics.

By means of the notion of generic identity, I am also able to draw a further distinction among FORs, which is crucial for understanding formal ontological fundamentality. This involves distinguishing *simple FORs* from *complex FORs*. Simple

FORs are FORs that are generically identical *only* to themselves. Primitive numerical identity is a plausible example of a simple FOR. Complex FORs, by contrast, are generically identical to some generically different FORs jointly. A good theoretical example of such a relation is Lowe's exemplification discussed above.

The distinctions between simple and complex FORs and basic and derived FORs crosscut. Every simple FOR is basic (hence proto formal ontological) because it is generically identical only to itself. By contrast, not every basic FOR is simple; there hold both basic and derived complex FORs. The reason for this is simple: there can obtain a proto FOR whose holding between a_1, …, a_n is not generically identical to the joint holding of some generically different proto FORs holding between entities *some of which are distinct* from a_1, …, a_n. Rather, its holding between a_1, …, a_n is generically identical to the joint holding of generically different proto FORs holding between a_1, …, a_n: it is basic. So, I distinguish between simple and complex basic FORs on the one hand and between basic and derived complex FORs on the other.

2.2 Formal Ontological Fundamentality and Non-Fundamentality

In the previous section, I argued that for an entity to have an ontological form is for it to be a relatum of a FOR or FORs jointly in an order. Consequently, for an entity to have a *simple ontological form* is for it to be a relatum of a simple FOR in an order, and for it to have a complex ontological form is for it to be a relatum of a complex FOR in an order. For example, on the assumption that an entity is primitively numerically identical, it has the simple ontological form of being numerically identical. Primitive numerical identity is a simple FOR since it is basic and its holding is generically identical only to itself.

Simple ontological forms are *fundamental ontological forms* because simple FORs are *fundamental FORs*. Simple FORs are basic and their holding does not consist of anything: their holding is generically identical only to themselves. Formal ontological fundamentality is being unconstituted in the sense of generic identity. Thus, for an entity to have a fundamental ontological form is for it to be a relatum of a simple

FOR in an order.[8] *The full fundamental ontological form* of an entity is generically identical to a simple FOR or FORs jointly in an order.[9]

I can illustrate this with primitive numerical identity again: primitive numerical identity is a fundamental ontological form of the identical entity, but not necessarily its full fundamental ontological form. Primitive substances in Lowe, for instance, bear other simple FORs than numerical identity to some entities, such as instantiation (1998, 169–73).

The full fundamental ontological form does not have to be the full mere ontological form either. An entity may have the full fundamental ontological form and bear a *derived* FOR to something. For example, an entity *can* be fundamentally a part and numerically identical but bear the derived FOR of proper parthood to a whole that has two proper parts.[10] In that case, being a relatum of the simple FORs of parthood and numerical identity does not exhaust the ontological form of this part: its ontological form is partly generically identical to the holding of the derived FOR of proper parthood. Therefore, I should say that the *full ontological form* of an entity is generically identical to a FOR or FORs jointly in an order.

If the full ontological form of an entity involves a derived FOR (proto or not), the full ontological form is *non-fundamental*. Hence, for an entity to have a *non-fundamental ontological form* is for it to be a relatum of a derived FOR in an order. Any derived FOR is a non-fundamental FOR for the very reason that its holding is generically identical to the joint holding of generically different FORs that involve additional necessary relata entities. The sole existence of the relata of the derived FOR is not exhaustively necessary for the holding of the derived FOR – even if their existence was jointly suffi-

[8] Formal ontological fundamentality is not to be confused with the fundamentality of entities, which may be dubbed "ontological fundamentality". As such, the possible fundamental ontological form of being a property, for example, differs from the putative fact that some but not all properties are fundamental entities (Tahko 2018).

[9] Note that it might be possible that there is no fundamental ontological form since no simple FORs hold. "Gunky" formal ontology in which every ontological form is complex seems to be possible.

[10] The example denies holism because the part has primitive numerical identity and it could exist without the whole.

cient. The holding of every derived FOR is constituted in terms of generic identity. The generically different FORs are then *individually* more fundamental and the ontological form of the relata of the derived FOR is at least partly generically identical to them. Therefore, the *fundamental* form of existence of these *relata* (if they have one) – that is, the fundamental way in which they exist – cannot be even partly generically identical to the holding of the derived FOR.

To use the previous example of the whole with two parts again, the proper parthood of the first part is its non-fundamental ontological form. The mere existence of this part and the whole is not exhaustively necessary for the holding of proper parthood between them, since the *whole* would not exist and proper parthood would not hold without the existence of the other part of the whole. So, the holding of proper parthood between the first part and the whole is somehow derived from FORs that have the other part as a relatum (proper parthood is standardly considered dyadic). For the first part, bearing proper parthood to the whole is not something that this part fundamentally is in formal ontological terms. By contrast, the whole would not have *any* ontological form without the non-fundamental ontological form of bearing proper parthood to two numerically distinct entities. The whole would not exist without the two parts existing. This whole is a non-fundamental entity in formal ontological terms (bearing in mind the anti-holism of the example).

3. Fundamental Ontological Form of Tropes

Now I can apply my metatheory about formal ontology of the previous sections to the SNT of tropes, substances and the relation of inherence between them.[11] *Ontologically*, tropes are entities that are standardly identified with characters or natures – that is, what tropes are like. Plausible examples of tropes or characters in scientifically informed metaphysics are determinate basic quantities: rest masses, charges and spin quantum numbers. These characters are “thin” or qualitatively simple: they do not even have aspects that would be numerically identical to them. They can also be indiscernible

[11] For the references to the articles presenting and defending the SNT, cf. the Introduction.

and numerically distinct; the ontological principle of the identity of indiscernibles does not hold true of tropes.

In this paper, however, my focus is on the *ontological form* of tropes – their (relational) form of existence rather than their ontology. According to the SNT, there are two *primitive* FORs *qua* terms: numerical identity and parthood. They are not defined in the theory. One of the *defined* FORs in it (i.e. FORs *qua* defined terms), numerical distinctness, is defined as the negation of numerical identity. Another defined FOR, proper parthood, is defined by numerical identity and parthood: x is a proper part of $y =_{\text{df}} x$ is a part of y AND x is not numerically identical to y.

The third defined ontological form is strong rigid (existential) dependence that is defined modally by the notion of existence, numerical identity and parthood. A contingent entity x is strongly rigidly dependent on a contingent entity y if and only if

(1) it is not metaphysically possible that x exists and y does not exist

(2) x and y are not numerically identical

(3) y is not a part of x (cf. Keinänen 2011, 431).

This differs from strong generic (existential) dependence that is the fourth defined ontological form in the SNT. Roughly, any contingent entity *x of kind P* is strongly generically dependent on some contingent entity *y of kind R* if and only if

(1) it is not metaphysically possible that any x of kind P exists and no y of kind R exists

(2) x of kind P and y of kind R are not numerically identical

(3) y of kind R is not a part of x of kind P (ibid.).

The numerical identity of tropes is not only a primitive term but also their *simple* reflexive FOR in the SNT. The holding of the FOR of numerical identity of each trope is not generically identical to any different FOR. Since every simple FOR is basic, the only explanation for the holding of numerical identity of each trope is the mere existence of the trope. This entails that each trope is a unity (i.e. one or countable) and an individual.

Another simple (and basic) reflexive FOR in the SNT is parthood, which holds of every trope: each trope is a part. Consider any trope whatsoever and its sole existence is sufficient and exhaustively necessary for it being a part. This ontological form distinguishes tropes from modes in Lowe's four-category ontology, for instance. Lowe denies that modes are parts (2006, 97). According to the SNT, however, no trope is a subject of the defined FOR of proper parthood (i.e. a whole). The SNT states that tropes are mereologically simple (mereological atoms).

The element that distinguishes the SNT from some other trope theories such as Campbell's (1990), Ehring's (2011, 98ff.) or Maurin's (2011) view is that tropes are strongly rigidly or generically *dependent* (cf. Simons 1994). It is not possible that there is a trope without some entity numerically and wholly distinct from the trope existing. The defined but simple (hence basic) FOR of strong rigid dependence or strong generic dependence holds of every trope. The standard case in the SNT is that there is a group of *mutually strongly rigidly* dependent tropes – that is, the nuclear tropes of a substance (more about this below).

In sum, according to the SNT the simple ontological forms of tropes are that they are strongly rigidly or generically dependent individual entities (i.e. numerically identical unities) that are simple parts. As I argued above, for an entity to have a fundamental ontological form is for it to be a relatum of a simple FOR in an order. Therefore, the holding of each of these simple FORs is a fundamental ontological form of any trope. Their joint holding is generically identical to the *full fundamental* ontological form of any trope. Fundamentally, in formal ontological terms, tropes are strongly rigidly or generically dependent individual simple parts. This is their fundamental character-neutral relational way of existence, which will turn out to be a crucial result for responding to certain objections to trope theory in the next section.

4. Applying the Fundamental Ontological Form of Tropes

Already from this, one can see that the dichotomy between objects and properties is not at all the right way to understand what tropes fundamentally are in formal ontological terms in the SNT, or any trope theory, *contra*, what Garcia and Maurin say about tropes, for example. For instance, if one follows Armstrong and understands his talk of tropes as junior substances meaning that tropes are fundamentally more akin to objects than properties, one is on the wrong track right from the start.

The formal ontological distinction between objects and properties presupposes some FOR holding between them. Objects are one relatum of this relation and properties are the other relatum. In realist metaphysical theories about universals, this relation is instantiation, participation or exemplification, depending on the theory. In nominalist theories that are committed to the existence of properties, it is for instance class/set membership, inherence, modification or characterization.

The SNT is among such theories, but it gives a reductive metaphysical *analysis* of the relation of inherence: inherence consists of other relations such as parthood and strong rigid dependence (cf. Keinänen's paper in this collection and Fisher 2018). This is due to the point that the SNT, like any trope theory, such as Maurin's account (2011), is a *bundle theory of objects*. For an entity to be an object is for it to be a terminus of inherence, since for an entity to be an object is for it to be a bearer of properties. In the bundle theories, objects are complex entities constructed by tropes (or, universals) and the holding of certain relations such as parthood and strong dependence or compresence, which are analysing relations (e.g. Campbell 1990, ch. 1; Maurin 2011; Fisher 2018). Therefore, not only objecthood but also inherence is in the *analysandum* in trope theories. So, it would be viciously circular if inherence was one the relations in the *analysans*. Consequently, in trope theories the holding of the analysing relations has to constitute the holding of inherence. Thus, inherence is not a simple FOR in the SNT, or any trope theory, because it is complex.

Hence, it is not stated in the SNT – as it ought not be in any trope theory – that tropes are *formal ontologically fundamentally* properties or objects, or more akin to one than the other. One just cannot consider the fundamental ontological form of tropes in these terms when one considers trope theories. Rather, ontologically tropes are identified with thin natures, and in the SNT their full fundamental ontological form is to be a strongly rigidly or generically dependent individual simple part. The dichotomous question set-up of properties or objects – or being more akin to one or the other – is a non-starter from the point of view of the SNT, as it should be in any trope theory as a bundle account of objects.

In order to argue against trope theory, Garcia (2016, 2) has recently introduced a distinction between module and modifier tropes. In terms of my metatheory of formal ontology, this distinction is based on the fundamental FOR of self-inherence. Module tropes are fundamentally self-inhering, whereas modifier tropes are not:

> In this stronger sense [of module tropes], 'particularizing a property' involves ascribing *objecthood* to a property. Here, particularization involves converting a shareable and singly characterizing property (an immanent [Armstrongian] universal) into a non-shareable and *thinly propertied object*: a module trope. So understood, the Slogan fixes on the concept of a module trope: a *primitively*, naturally, and thinly charactered *object*. (Ibid.; second, third and fourth emphases added)

> Here, the Slogan fixes on the concept of *a modifier trope*: a non-shareable and *non-self-exemplifying* property. (Ibid.; cf. 2016, 5; 2015, 138, 144, 148; emphases added)

So, take any F module trope and it just is F, but an F modifier trope is not F. Module tropes are fundamentally self-inhering thinly propertied objects while modifier tropes are not.

Garcia's distinction is also a non-starter in the SNT, or in any trope theory. This distinction presupposes that inherence can be a simple FOR holding of tropes because self-inherence is presumed to be such a relation. However, I argued above that as a trope theory, the SNT denies the formal ontological simplicity and hence fundamentality of inherence. Therefore, Garcia's distinction between module and modifier tropes does not apply to the SNT, or to any trope theory. Garcia's argument against trope theories, which is based on this distinction, does not hit the target at all (Garcia 2016; 2015; cf. Garcia 2014a, sec. 2).

In addition to properties or objects, another putative ontological form typically associated with tropes is that they are *particulars* in contrast to universal thin natures. This is not correct about the ontological form of tropes in the SNT. Particularity is not among the ontological forms of tropes in it. Formal ontologically, particularity is not theoretically needed for anything: it does not do any theoretical work. There is no contrasting class here because as a nominalist theory, the SNT is not committed to the existence of universals. It is only in discussions with the realists that we can inform them that tropes are the subjects but not the termini of the FOR of instantiation, participation or exemplification of the *realists*. However, the SNT cannot accommodate any of these FORs since each of them presupposes the existence of universals.

Some might object here that it does not presuppose universals to say that tropes are particulars because they can be indiscernible and numerically distinct. The principle of the identity of indiscernibles does not hold true of tropes. Therefore, even a nominalist can hold that tropes are particulars; Williams (1986) and Ehring (2011, 35), for instance, put particularity in this way.

My response to this possible objection is that my account of ontological form is not compatible with this characterization of the ontological form of particularity. *Being indiscernible from* (exactly resembling to many metaphysicians) is not a FOR since it is a character-dependent relation. The statement that *x* and *y* are indiscernible tells us, if true, something about the character of *x* and *y* even without further assumptions – namely that they are indiscernible and are of the same type or kind. Hence, when the formal ontological side of the SNT is put in terms of my metatheory of formal ontology, the SNT cannot accommodate Williams' and Ehring's characterization of particularity either. This does not mean, however, that it is false that the principle of the identity of indiscernibles does not hold true of tropes. On the contrary, the SNT is committed to the denial of this ontological principle, although it is not considered a formal ontological principle in the SNT.

If someone insisted here that this is a problem for my metatheory and that particularity is theoretically needed eventually, I could reply that for the nominalist, being a particular is like existence: among entities there is no contrast class. They are *maximally transcategorial*: an entity of any ontological form whatsoever is particular and exists. They cannot be the ontological forms of ontological forms either, since as non-entities, ontological forms do not have ontological forms. Therefore, being a particular or existing are not ontological forms in nominalism; rather, they fall under ontology (given one insists on particularity).

My metatheory also undermines any argument against tropes that is premised upon the holding of indiscernibility, exact resemblance/similarity, resemblance/similarity or their opposite *in respect of some ontological form*. For instance, one might claim that the nuclear tropes of a simple substance in the SNT are exactly resembling with respect to particularity or strong rigid dependence. Ehring (2011, 182) and Garcia (2014b, secs. II–IV) present examples of such arguments in the literature.

One of the premises of Ehring's argument is that there are tropes that are exactly or inexactly similar "with respect to their particularity" (2011, 182). Garcia's argument is based on the notion of dependency profile: "A trope *t*'s dependency profile specifies all the distinct token and/or types of tropes on which *t* is (rigidly or generically) dependent" (2014b, 169). If the dependency profile of a trope were in its character, then the SNT would deny it. According to the SNT, tropes do not depend for their existence because of their character. So, the charitable reading of Garcia is that in the SNT, the dependency profile of a trope has to consist of the FORs of strong rigid and generic dependence that this trope bears. Of the dependency profiles, Garcia maintains that they "admit of qualitative differences and similarities" and goes on to argue against the SNT with that claim (ibid. 170 and secs. III & IV). Garcia's argument against the SNT is then premised upon the statement that the FORs of strong rigid and generic dependence can stand in the relations of similarity and difference.

Let us grant to Ehring for the sake of the argument that tropes have the ontological form of particularity, even though I could rebut his argument simply by saying that the SNT does not have to state that tropes are particulars. Proceeding with this assumption, I can respond to Ehring and Garcia that according to my metatheory, *no entity bears indiscernibility, exact resemblance/similarity, resemblance/similarity or their opposite in respect of any ontological form*. The reason for this is simple: these character-dependent relations can hold only among entities. However, ontological forms – that is, FORs – are not entities in themselves; FORs are internal relations. Thus, ontological forms or FORs can stand in neither indiscernibility, exactly resemblance/similarity, resemblance/similarity nor their opposite. This undermines any argument that assumes such a standing, especially Ehring's and Garcia's lines of reasoning that take the putative ontological forms of the particularity and strong rigid or generic dependence of tropes as their targets.

Ehring (2011, 179–80) has another argument against standard tropes that are simple entities identified with thin natures or characters: they are not simple, *pace* the SNT. This argument is similar to Herbert Hochberg's earlier line of reasoning (2004, 39; cf. Moreland 2001, 70–1; Armstrong 2005, 310). A key premise in Ehring's argument is that "arbitrarily different internal relations" must have distinct relata. The holdings of arbitrarily different internal relations vary or are realized independently of each other. Exact resemblance and numerical distinctness among standard tropes are arbitrarily different internal relations. Thus, they have distinct relata and no individual trope can be both exactly similar, numerically distinct and simple. The simplicity of standard tropes is refuted (Ehring 2011, 177–80).

Hochberg's (2004, 39) key premise is that the internal relations of exact similarity and numerical distinctness *qua* logically independent basic propositions cannot have the same truthmakers. If it is logically possible that any basic proposition is true and another false (and *vice versa*), then these basic propositions are logically independent.

My metatheory supplies the SNT with resources to answer these arguments (cf. Hakkarainen & Keinänen 2017). The SNT can deny both Ehring's and Hochberg's key premises and therefore refute their arguments against simple tropes. Exact resemblance is a character-dependent internal relation, whereas numerical distinctness is a character-neutral internal relation – that is, a FOR. Yet their holdings can have grounds that are not numerically distinct. Let us assume that there are two numerically distinct exactly resembling tropes. As I argued above, the ground of their numerical distinctness is nothing but their existence. Ontologically, each of these two tropes is identified with a character. So, there holds no numerical distinctness between their existence and character; they are entities identified with the character that they are. Now, these two tropes exactly resemble because of the characters they are. Thus, the grounds of these tropes being numerically distinct and exactly resembling are not numerically distinct. These two internal relations do not have to have numerically distinct relata or truthmakers *qua* propositions.

Nonetheless, numerical distinctness and exact resemblance among tropes in general can be arbitrarily different internal relations, or their propositions can be logically independent. Let us consider the former first. The holding of any exact resemblance depends on the character of its relata. By contrast, the holding of the FOR of numerical distinctness between tropes in general does not depend on the characters that the tropes are. So, tropes may or may not be numerically distinct independent of the characters they are. Tropes of exactly similar or different character can be numerically distinct. The holdings of numerical distinctness and exact resemblance can vary or be realized independently from one another, which is a sufficient condition for them being arbitrarily different. Furthermore, this independence of variation and realization may also be construed as *logical* in nature. Thus, numerical distinctness and exact resemblance among simple tropes are also logically independent as propositions. Hence, they are both logically independent *qua* basic propositions and arbitrarily different internal relations. As was seen just above, these logically independent basic propositions do not have to have distinct truthmakers or arbitrarily different internal relations distinct relata. Thus, the SNT denies Ehring's and Hochberg's key premises and hence refutes their arguments.

5. Of the Non-Fundamental Ontological Form of Tropes

According to the SNT, tropes also have *derived* ontological forms. Let us take two examples, although the main topic of this paper is not the non-fundamental ontological forms of tropes. In the SNT, every trope is necessarily a *proper part* of a substance: there are no "free-floating tropes". To expound the constitution of the relation of proper parthood between tropes and substances, let us first facilitate the presentation and take the example of an arbitrary minimal substance in the SNT. Such a substance is simple, since it does not have parts that are substances; it has only two trope parts – say, a rest mass trope and a charge trope. It is a minimal substance. Let us also assume that the two tropes are mutually strongly rigidly dependent: neither of them can exist without the other. They are the only tropes and nuclear tropes of the minimal substance. Consequently, also their plurality has to exist given the contingent existence of one of them.

By the Conditioning Principle adapted from Simons (1987, 322), the plurality of these two tropes is not existentially dependent on any other entity than the rest mass trope and the charge trope. This principle states that necessarily, if plurality *x* is such that every dependent entity of it (its element) has all the entities on which the entity depends also in *x*, then *x* is not dependent on anything else than its elements. In other words, the elements of *x* satisfy its "existential needs". Thus, the plurality of the two tropes is strongly rigidly *independent*: it does not depend for its existence on any entity that is wholly distinct from it – that is, does not share parts with it. The plurality depends for its existence only on the rest mass trope and the charge trope. Since the definition of strong rigid dependence rules out dependence on parts, the plurality satisfies the condition of being a minimal substance in the SNT: it is strongly rigidly independent. Hence, here we actually have an individual.

Regarding the relation of proper parthood holding between the arbitrary minimal substance and its two mutually rigidly dependent trope parts, the upshot is that the *holding* of this relation from one of the tropes to the substance requires the existence of the other trope. This result generalizes in the

SNT. Thus, proper parthood from any trope to a simple substance is a *derived* FOR (note that it is standard to consider proper parthood dyadic).

It follows that this FOR of proper parthood between a trope and a simple substance is neither a simple nor a fundamental ontological form of any trope in the SNT, in contrast to the parthood reflexively holding of tropes. Since the trope is an arbitrary trope in the SNT, no trope is *fundamentally* a proper part of a substance. Proper parthood between them is neither a simple nor a basic FOR.[12]

Connected to this, recall that Armstrong claims that tropes are junior substances. If this involves that tropes have the ontological form of independence, it is not correct in the SNT either. Tropes are strongly rigidly or generically *dependent* entities, whereas even simple substances are strongly rigidly *independent* entities.

Concreteness is another derived ontological form of each trope. According to the SNT, every trope is located in space-time. This entails that no trope is abstract – that is, an entity not having even a temporal location. Assuming that concreteness is an ontological form, some FOR has to hold between each trope and space-time. What this FOR is depends on the theory of space-time. Yet it must be a *derived* FOR because the mere existence of an arbitrary trope and space-time is not exhaustively necessary for its holding (it may be sufficient though). At least the existence of a relational trope or space-time point is required. Hence, the ontological form of tropes being concrete is derived.

The derived status of concreteness and the proper parthood of a substance does not mean, however, that no trope is *necessarily* concrete and a proper part of a substance. Rather, the SNT states that every trope is necessarily a proper part of some simple substance. Let us take nuclear tropes as an example (bracketing the limiting case of singular nuclear tropes). Necessarily, if there is an arbitrary nuclear trope, then there is another trope or there are other tropes and these tropes are strongly rigidly dependent on each other. By the Conditioning Principle, it follows that the arbitrary trope is

[12] This also means that being a substance is not a fundamental ontological form in the SNT.

also necessarily a proper part of a simple substance. Equally, necessarily for any arbitrary trope, the trope exists in a spatio-temporal location. Proper parthood between an arbitrary nuclear trope and a simple substance and the concreteness of every trope are derived *proto* FORs in the SNT. It is necessary to any trope that it is a relatum of these two FORs.

6. Conclusion

I have argued that according to the Strong Nuclear Theory (SNT), the full fundamental ontological form of every trope is to be a strongly rigidly or generically dependent individual entity that is a simple part. In these formal ontological terms, each trope is concrete and non-fundamentally but necessarily a proper part of a simple substance. The proper parthood and concreteness of every trope is one of its derived ontological forms. Ontologically, the SNT identifies each trope with a thin character. It is also rather an ontological than a formal ontological feature of each trope in the SNT that the principle of the identity of indiscernibles does not hold true of it.

This summarizes the way in which I put the SNT in terms of my metatheory of formal ontology and its difference from ontology. Ontology studies questions of existence, such as whether there are properties from the unique point of view provided by formal ontology. The core subject matter of formal ontology is ontological form, of which I have a relational account. My account employs the notion of generic identity, which is a form of generalized identity distinguished from familiar numerical or objectual identity.

In terms of generic identity, for an entity to have an ontological form is for it to be a relatum of a formal ontological relation or relations jointly – that is, a character-neutral internal relation or relations jointly in an order. Internal relations really hold of their relata, although they are not entities numerically distinct from the relata. The holding of internal relations can, in principle, be asserted by true relational statements. The statements about character-neutral internal relations do not say anything about the character of the relata of the relations without further assumptions (*contra* character-dependent internal relations). The character of an entity is what the entity is like. In addition to entities themselves, their possible characters belong to the extension of the concept of existence or being, which I assume to be interchangeable univocal concepts.

In the fourth section, I showed that putting the SNT in terms of my metatheory is fruitful because it gives resources to answer the arguments against tropes advanced by Douglas Ehring, Robert K. Garcia and Herbert Hochberg. These arguments do not distinguish the formal ontology of tropes from their ontology. Their argumentative gap lies in overlooking this distinction.

I distinguish the fundamental ontological form from the non-fundamental by the distinction between simple and derived formal ontological relations that builds upon a tripartite distinction among proto, derived and basic internal relations. For an entity to have a fundamental ontological form is for it to be a relatum of a simple formal ontological relation in an order. Every simple formal ontological relation is basic. The mere existence of the relata of a simple formal ontological relation is jointly sufficient and exhaustively necessary for the holding of the relation. It is also simple because the holding of such a relation is generically identical only to itself. Thus, the holding of a simple formal ontological relation is not constituted in the sense of generic identity.

By contrast, for an entity to have a non-fundamental ontological form is for it to be a relatum of a derived formal ontological relation in an order. The holding of a derived formal ontological relation is, roughly, generically identical to generically different formal ontological relations that jointly hold of entities some of which are numerically distinct from the relata of the derived formal ontological relation.

Applying my metatheory to the SNT and trope theories establishes that the typical dichotomous set-up of asking whether tropes are fundamentally properties rather than objects is a non-starter in the SNT. The same is correct of talk about tropes as particular properties or particularized properties when one understands this talk to concern the fundamental ontological form of tropes. First, the SNT does not need particularity theoretically for anything. Secondly, being a property, object or something akin to one or the other presupposes the relation of inherence that is not a fundamental formal ontological relation in the SNT or any trope theory as a bundle theory of objects.

University of Tampere

References

Allen, Sophie R. (2016), *A Critical Introduction to Properties*, Bloomsbury, London.

Armstrong, D. M. (1989), *Universals*, Westview Press, Boulder.

Armstrong, D.M. (2005), "Four Disputes About Properties", *Synthese*, 144, pp. 309–320.

Campbell, K. (1990), *Abstract Particulars*, Blackwell, Oxford.

Correia, F. (2017), "Real Definitions", *Philosophical Issues* 27, pp. 52–73.

Correia, F. & A. Skiles. (2017), "Grounding, Essence, And Identity", *Philosophy and Phenomenological Research*: doi: 10.1111/phpr.12468.

Divers, J. (2002), *Possible Worlds*, Routledge, London.

Dorr, C. (2016), "To Be F is To Be G", *Philosophical Perspectives* 30, pp. 39–134.

Edwards, D. (2014), *Properties*, Polity Press, Cambridge.

Effingham, N.; Beebee, H. & Goff, P (2010), *Metaphysics: The Key Concepts*, Routledge, London.

Ehring, D. (2011), *Tropes: Properties, Objects, and Mental Causation*, Oxford University Press, Oxford.

Fisher, A. R. J. (2018), "Instantiation in Trope Theory", *American Philosophical Quarterly* 55, pp. 153–164.

Garcia, R. (2014a), "Bundle Theory's Black Box: Gap Challenges for the Bundle Theory of Substance", *Philosophia* 42, pp. 115–126.

Garcia, R. (2014b), "Tropes and Dependency Profiles: Problems for the Nuclear Theory of Substance", *American Philosophical Quarterly* 51, pp. 167–176.

Garcia, R. (2015), "Is Trope Theory a Divided House?", in Galluzzo, G. & Loux, M. (eds.), *The Problem of Universals in Contemporary Philosophy*, Cambridge University Press, Cambridge, pp. 133–155.

Garcia, R. (2016), "Tropes as Character-Grounders", *Australasian Journal of Philosophy* 94, pp. 499–515.

Hakkarainen, J. & Keinänen, M. (2016), "Bradley's *Reductio* of Relations and Formal Ontological Relations". in Laiho, H. & A. Repo (eds.), *DE NATURA RERUM - Scripta in honorem professoris Olli Koistinen sexagesimum. Reports from the Department of Philosophy Vol. 38.* University of Turku, Turku, pp. 246–261.

Hakkarainen, J. & Keinänen, M. (2017), "The Ontological Form of Tropes – Refuting Douglas Ehring's Main Argument against Standard Trope Nominalism", *Philosophia*, 45 (2), pp. 647–658.

Hakkarainen, J., Keinänen, M. & Keskinen, A. (2018), "Taxonomy of Relations", in Bertini, D. & D. Migliorini, (eds.), *Relations. Ontology and Philosophy of Religion*. Mimesis International, Verona, pp. 93–108.

Hochberg, H. (2004), "Relations, Properties and Predicates", in H. Hochberg and K. Mulligan (eds.), *Relations and Predicates*, Ontos Verlag, Heusenstamm, pp. 17–53.

Keinänen, M. (2011), "Tropes—The Basic Constituents of Powerful Particulars?", *Dialectica* 65, pp. 419–450.

Keinänen, M. & Hakkarainen, J. (2010), "Persistence of Simple Substances", *Metaphysica* 11 (2), pp. 119–135.

Keinänen, M. & Hakkarainen, J. (2014), "The Problem of Trope Individuation", *Erkenntnis* 79 (1), pp. 65–79.

Keinänen, M., Keskinen, A & Hakkarainen, J. (2017), "Quantity Tropes and Internal Relations", *Erkenntnis*: doi.org/10.1007/s10670-017-9969-0.

Linnebo, Ø. (2014), "'Just is'-Statements as Generalized Identities", *Inquiry* 57, pp. 466–482.

Lowe, E. J. (1998), *The Possibility of Metaphysics*, Oxford University Press, Oxford.

Lowe, E. J. (2006), *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Clarendon Press, Oxford.

Lowe E. J. (2012), "A Neo-Aristotelian Substance Ontology", in T. E. Tahko (ed.), *Contemporary Aristotelian Metaphysics*, Cambridge University Press, Cambridge, pp. 229–248 .

Maurin, A.-S. (2011), "An Argument for the Existence of Tropes", *Erkenntnis* 74, pp. 69–79.

Maurin, A.-S. (2018), "Tropes", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), URL = <https://plato.stanford.edu/archives/sum2018/entries/tropes/>, retrieved 12/9/2018.

Moreland, J. P. (2001), *Universals*, McGill-Queen's University Press, Montreal & Kingston.

Mulligan, K. (1998), "Relations: Through Thick and Thin", *Erkenntnis* 48, pp. 325–353.

Rayo, A. (2013), *The Construction of Logical Space*, Oxford University Press, Oxford.

Simons, P. (1994), "Particulars in Particular Clothing: Three Trope Theories of Substance", *Philosophy and Phenomenological Research* 54, pp. 553–575.

Simons, P. (1987), *Parts: A Study in Ontology*, Oxford University Press, Oxford.

Smith, B. (1978), "An Essay in Formal Ontology", *Grazer Philosophische Studien* 6, pp. 39–62.

Smith, B. (1981), "Logic, Form and Matter", *Aristotelian Society Supplementary Volume* 55, pp. 47–74.

Smith, B. & Mulligan, K. (1983), "Framework for Formal Ontology", *Topoi* 2, pp. 73–85.

Smith, B. & Grenon, P. (2004), "The Cornucopia of Formal-Ontological Relations", *Dialectica* 58, pp. 279–296.

Tahko, T. E. & Lowe, E. J. (2015), "Ontological Dependence", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), URL = <http://plato.stanford.edu/archives/spr2015/entries/dependence-ontological/>, retrieved 23/11/2016.

Tahko, Tuomas E. (forthcoming), "Fundamentality", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), URL = <http://plato.stanford.edu/archives/fall2018/entries/fundamentality/>, retrieved 12/9/2018.

Van Inwagen, P. (2009), "Being, Existence and Ontological Commitment", in Chalmers, Manley and Wasserman (eds.), *Metametaphysics. New Essays on the Foundations of Ontology*, Clarendon Press, Oxford, pp. 472–506.

Williams, D.C. (1986), "Universals and Existents," *Australasian Journal of Philosophy* 64 (1), pp. 1–14.

A Trope Theoretical Analysis of Relational Inherence

MARKKU KEINÄNEN

1. Introduction

Trope bundle theories of substance (e.g. Williams 1953; Campbell 1990; Maurin 2002; Keinänen 2011, Keinänen & Hakkarainen 2010, 2014; Giberman 2014) aim to construct objects and all other entities by means of aggregates of tropes. Tropes are thin particular natures like a particular *–e* charge or a particular roundness in some location. Thus, tropes are themselves concrete in the sense of having some specific spatial or spatio-temporal location. In trope theories, objects and all other particulars are constructed as mereological sums of tropes that fulfil certain conditions. For instance, objects are identified with mereological sums of mutually co-located ("concurrent" or "compresent") tropes (cf. Williams 1953; Campbell 1990). The thin nature of a trope is contrasted with the thick nature of the object constituted by distinct tropes.[1]

It has been customary to consider tropes as *particularized qualities* or *particular properties of objects* (cf. Armstrong 1989, Allen 2016). The standard ways to pick out and identify tropes as properties of objects (like the redness of some rose) have provided support to this intuitive conception.[2] Nevertheless, being a particular property is a primitive category

[1] In Williams' and Campbell's classical trope theories, tropes are also considered as "abstract" in the sense of having capability of being co-located with other tropes, Fisher (2018, sec.1).

[2] As does Lowe (2003), we have distinguished between individuation in the epistemic sense (i.e., picking out an entity in our thought) and individuation in the metaphysical sense. (i.e., the determination of the identity conditions of an entity). Moreover, we defend the idea that tropes have primitive identity conditions, Keinänen & Hakkarainen (2014).

feature of *modes*, which primitively inhere in (or, characterize) objects.[3] By contrast, most trope theories (i.e. trope bundle theories) aspire to *analyze* monadic inherence (objects having tropes), e.g., by means of parthood, co-location and/or existential dependencies.[4] Therefore, being a property (or, being an object) is not left primitive.

The trope theoretical analysis of monadic inherence can be regarded as a case of *metaphysical reduction*: in the analysis of inherence, a central feature of reality (objects having properties) is reduced to the holding of a fact about the basic entities of the category system (tropes). In the basic level, there are assumed to be only tropes that form objects if the respective aggregate of tropes fulfils certain conditions.[5] Correspondingly, the object has a trope as its property if and only if it has the trope as a certain kind of part. We may take a trope theory that identifies objects with mereological sums of co-located tropes as a simple example. Assume that object *i* is a mereological sum of three mutually co-located tropes t_1, t_2 and t_3, which are determinate quantities. Let t_1 be a determinate –e charge, t_2 a determinate mass, and t_3 a determinate spin quantum number. Object *i* has trope t_1 as its property (i.e., *i* has a –e charge) if and only if *i* has trope t_1 as its part and t_1 is co-located with *i*. Thus, in the trope theoretical analysis of inherence, the reduction is assumed to take place in the general level of ontological categories: the facts about objects and properties are assumed to be identified with the facts about tropes and the complex entities tropes form.

In trope theories, the "traditional" object-property dichotomy is explained away. Neither of these two categories – *objects* (entities characterized by properties) or *properties*

[3] Modes are introduced by the different substance attribute theorists in a two category ontology of substances and modes (cf. Martin 1980; Heil 2012), or in Neo-Aristotelian four-category (Lowe 2006, 2009, 2015) and six-category (Ellis 2001) ontologies.

[4] For instance, classical trope theories (Williams 1953, Campbell 1990) analyze monadic inherence in terms of parthood and co-location: trope *t* is a property of object *i* if and only if *t* is a part of *i* and *t* is co-located with *i*.

[5] According to Williams (1953) and Campbell (1990), tropes are existentially independent entities and objects are mereological sums of co-located (compresent, concurrent) tropes. Cf. Keinänen (2011, sec. 3) and Fisher (2018) for further discussion.

(entities inhering in or characterizing objects) – retains its status as a basic category. Fundamentally, tropes are neither properties nor objects. Although tropes are conveniently identified (or picked out) as "properties of their bearers" (like e charge of a positron or redness of a rose), they are particular natures – entities of a single fundamental category – which constitute all (or almost all) other entities.

Nevertheless, the trope theoretical analysis of inherence remains silent about relations or relational entities. Dealing with the question about the existence and ontological status of relations has turned out to be difficult for trope theorists. Most trope theorists have recently held either of the two main alternative views about relations, which, as I will argue, are both unsatisfactory. The first is the *eliminativist* view adopted by Keith Campbell (1990) and Peter Simons (2014, 2016), according to which there are no relations or relational entities. Everything that exists is constituted by monadic tropes. Tropes are connected by different kinds of internal relations, but internal relations are not relational entities additional to their relata.[6] Secondly, the advocates of the *relata specific v*iew - Anna-Sofia Maurin (2002, 2010, 2011), Jan-Willem Wieland and Arianna Betti (Wieland & Betti 2008; Betti 2015) - introduce relational tropes in addition to monadic tropes.[7] The existence of relational trope *r* is assumed to entail that *r* relates (or, relationally inheres in) certain specific relata *a* and *b*.

As I will argue in section 2, the relata specific view is unsatisfactory because it re-introduces the primitive dichotomy between characterizing (relations) and characterized entities (objects) at the level of relations. According to the relata-specific view, there are both primitively relating (relational tropes) and primitively related entities (objects). The relata-specific view leaves *relational inherence* as a primitive formal ontological relation between relational tropes and their relata. Thus, a trope theorist adopting the relata specific view loses

[6] For different kinds of internal relations, cf. Keinänen, Keskinen & Hakkarainen (2017, sec.2).

[7] Officially, Wieland & Betti (2008) and Betti (2015) stay neutral between tropes and modes. Moreover, they allow for the possibility of relata specific relation universals. However, they work out their position by considering the relata specific entities as tropes.

one of the main benefits of trope theories, which is the general analysis of inherence. In order to retain the initial attraction of trope theories, eliminativism might seem to be an appealing option. In section 3, I argue that eliminativism sets serious limitations to the ontological explanatory power of trope theories. In addition to spatio-temporal relations, the current scientific theories have introduced entities which are serious candidates for entities to be best categorized as relations or relation-like existents.

Therefore, the main objective of section 4 is to present a new trope theoretical analysis of relational inherence. The aim is to offer a metaphysical reduction of relational inherence, that trope *r* relates two or more entities. In other words, I reduce the holding of relational inherence to the obtaining of certain other relations in the trope theoretical category system. The analysis generalizes the trope theoretical analysis of inherence provided by the Strong Nuclear theory (SNT) (Keinänen 2011; Keinänen & Hakkarainen 2010, 2014) to relation-like tropes, r-tropes, for short. Section 5 deals with asymmetric and non-symmetric relations, which are a prima facie difficult case for the analysis, by assuming that all fundamental relations are quantities. Finally, in section 6, I provide a completely new account of the location of r-tropes.

2. The relata specific view

Anna-Sofia Maurin (2002, 2010, 2011), Jan-Willem Wieland & Arianna Betti (2008) and Betti (2015) have recently made an important contribution to trope ontology by defending relational tropes.[8] According to their view, relational tropes are primitively relating and relata specific entities. Assume that trope *r* is a relata-specific relational trope of 1 m distance between two objects *a* and *b*. Although there are minor differences in the different formulations of the relata specific view, a relata specific relational trope *r* is assumed to fulfil the following three conditions:

[8] However, Betti (2015, 100ff.) considers her defense of the relata specific view conditional: if we must introduce relations at all, the relata specific view constitutes the best account of relations.

1. Necessarily, if relational trope *r* exists, its relata, *a* and *b*, also exist. To put this in formal ontological terms, trope *r* is multiply rigidly dependent (only) on its relata, *a* and *b*.[9]

2. Necessarily, if trope *r* exists, *r* relates (i.e., relationally inheres in) *some* relata.

3. Necessarily, if trope *r* exists, *r* relates (relationally inheres in) its specific relata, *a* and *b*.

Thus, according to the relata specific view of relational tropes (henceforth, *the relata specific view*), necessarily, if trope *r* exists, objects *a* and *b* are in a 1 m distance from each other. In other words, the sole existence of a relational trope is considered to entail that certain relational fact obtains.

The relata specific relational tropes are introduced in order to avoid the *modal version of Bradley's regress*, in which the condition that starts the regress is formulated in modal terms. Assume that relational trope of 1 m distance *r* and its relata *a* and *b* exist. The general worry in this version of Bradley's regress is that, *prima facie*, the existence of an external relation and its relata does not entail that the relation relates its relata. The postulation of additional relations – such as the relation of instantiation connecting the relation and its relata – would only transfer the problem to a higher level (Wieland & Betti 2008; Maurin 2010, 2011). Relational tropes seem to solve the regress problem because the existence of certain relational trope *r* already entails that the relation between specific objects holds. For instance, the existence of 1 m distance trope *r* entails that objects *a* and *b* are in 1 m distance from each other. Because the existence of *a* and *b* does not entail the existence of *r*, the distance relation is contingent and external to its relata, objects *a* and *b* (Wieland & Betti 2008, sec.3).

Returning to what the relata specific view entails, conditions 1-3 are not independent of each other. It is fairly easy to observe that if relational trope *r* fulfils condition 3, it also sat-

[9] Let "≤ " be a relation of improper parthood between entities and "E!" the predicate of (singular) existence. "SRD (e,f)" = e is strongly rigidly dependent on f. The multiple rigid dependence of t on f and g, "MRD (t, (f, g)", can be presented as follows: MRD (t, (f, g)) = $\Box(E!t \rightarrow (E!f \wedge E!g \wedge \neg(f \leq t) \wedge \neg(g \leq t) \wedge \neg(f \leq g) \wedge \neg(g \leq f))) \wedge \neg (\Box E!f) \wedge \neg (\Box E!g) \wedge \neg (SRD(f, g)) \wedge \neg (SRD (g, f))$.

isfies the two first conditions. Trivially, if trope *r* relates certain specific relata, trope *r* relates some relata (3 entails 2). Moreover, the holding of a relation between specific relata entails that the relata exist. Therefore, assuming that *r* is a relata specific relational trope – that the existence of *r* is sufficient to its relating objects *a* and *b* – entails that *r* is also multiply rigidly dependent on *a* and *b* (3 entails 1). However, the converse does not hold (1 does not entail 3): multiple rigid dependence of trope *s* on two entities *a* and *b* does not entail that *s* relates these two entities.[10]

The last-mentioned point requires some discussion because there has been confusion about the role of multiple rigid dependence in the metaphysical explanation of relata specificity. Multiple rigid dependence (MRD, for short) is a *formal ontological relation* that connects mereologically disjoint contingent existents. MRD spells out how its relata can exist as the constituents of the world. However, the constraints MRD sets on its relata are minimal: necessarily, if entity *s* exists (somewhere, somewhen), then its dependees *a* and *b* also exist (somewhere, somewhen). In addition to holding between a relational trope and its relata, MRD can hold between events and the specific objects involved in these events or between borders and the objects confined by these borders, for instance. In order to distinguish between different kinds of entities which are multiply rigidly dependent on some other entities (e.g., between borders and relational tropes), we are obliged to provide a more detailed description of their category features.

According to the relata specific view, it is a primitive category feature of relational tropes that they relationally inhere in (i.e., relate) certain specific relata. In other words, *specific relational inherence* is not analyzed further and it is supposed to be a primitive formal ontological relation connecting its relata.[11] Specific relational inherence fixes the categorial na-

[10] As MacBride (2011, 173) observes, "[n]ecessary coexistence of a relation and its terms is not enough to ensure that the relation holds between its terms". To be more exact, the holding of 1 does not guarantee that *r* is a relational trope and that 3 holds.

[11] The most explicit advocates of the relata specific view, Wieland & Betti (2008) do not directly characterize specific relational inherence ("relating specific entities") as a formal ontological relation. However, they assume

ture of relational tropes as a specific kind of relational accident (e.g., in contradistinction to borders and events). This formal ontological relation spells out what general kind of entities relational tropes are and how they can exist as constituents of the world.[12] The primitive relational inherence is comparable to the formal relation of characterization ("monadic inherence") between modes (particular properties) and objects E.J. Lowe (2006, 2009, 2015) introduces in his Four-Category Ontology (cf. Keinänen 2018, sec.3). Like characterization, specific relational inherence is considered to be an internal relation: necessarily, if given entities occurring in specific relational inherence (a relational trope and its relata) exist, specific relational inherence holds between its relata. The advocates of the relata specific view claim to avoid Bradley's relation regress by assuming that the existence of the entities connected by specific relational inherence is sufficient to the holding of specific relational inherence.[13]

The relata specific view faces three serious difficulties. The first is that the relata specific view introduces particular relations (i.e. relational tropes) as a primitive ontological category. In other words, it introduces a distinction between primitively relating and primitively related entities. This distinction is parallel to the primitive distinction between modes and objects (particular attributes and substances).[14] One of the central motivations of trope theory is to eliminate the substance attribute distinction by means of the analysis of inherence (cf. Campbell 1990, secs. 1.1-1.6). If trope theorists must re-introduce a parallel distinction in the case of relational tropes, this seriously reduces the attraction of trope theories.

Moreover, there are two more specific problems, which are closely connected to the first. The first of these problems is a consequence of the fact that specific relational inherence en-

it to be a part of the nature of relational tropes (as entities belonging to a certain category) that they relate certain specific relata (ibid, sec. 3).

[12] Cf. Hakkarainen & Keinänen (2017) and Hakkarainen (2018) for more on formal ontological relations.

[13] By using our terminology (cf. Keinänen, Keskinen & Hakkarainen 2017, sec. 2), specific relational inherence is assumed to be a *basic internal relatio*n between its relata.

[14] Particular properties or modes are recently advocated, e.g., by Lowe (2006, 2009, 2015), Ellis (2001) and Heil (2012).

tails multiple rigid dependence, but the converse does not hold. One can ask: can we find an analysis for relational inherence by means of multiple rigid dependence and some additional condition? Like the analysis of monadic inherence, such an analysis would reduce the number of the primitive formal ontological relations needed in trope theory. Moreover, one could bring much-needed clarity to the category system by analyzing relational inherence by means of more transparent primitive notions (such as parthood and rigid dependence). Since the relata specific view leaves relational inherence as a primitive formal ontological relation, the opportunity for a further trope theoretical clarification of the category system is lost in the case of relational inherence.

The third problem concerns the spatial or spatio-temporal location of relational tropes. Most trope theorists are inclined to adopt the ontological view that all entities are spatio-temporal particulars, which Peter Simons (2010, 207; 2016, 113) calls "naturalistic nominalism". Thus, let us assume that also relational tropes have a spatio-temporal (or, at least a temporal) location. Assume that *r* is a 1 m distance trope relating objects *a* and *b*. Trope *r* is determining the location of other entities, but it is difficult to determine the location of *r* (cf. Simons 2003, sec.2). The advocates of relational tropes have not provided any answer to this difficulty. This is unsatisfactory because relational inherence seems to entail restrictions to the location of relational tropes – for instance, that relational tropes are at least temporally co-located with their relata. Since we are unaware of the exact consequences of relational inherence, this casts doubt on using relational inherence as a basic notion of an ontological category system.

3. Eliminativism

According to eliminativism (Campbell (1990; Simons 2014, 2016), there are no relations – relational tropes or any other kind of relational entities. Thus, relations (or, relational tropes) are eliminated as a fundamental category.[15] The world

[15] Earlier, Simons (2003, 2010) postulated entities he called "relational tropes". Nevertheless, Simons' "relational tropes" are relational accidents, entities multiply rigidly dependent on two or more entities. He does not

is constituted by monadic tropes, which are particular natures.

Because of the serious problems of the relata specific view, the adoption of eliminativism with respect to relations might seem to be an attractive option for trope theorists.[16] Nevertheless, I will argue in this section that eliminativism is, if not provably false, at least a very risky position for two main reasons. The first is that eliminativism seriously restricts the available options in providing a trope theoretical account of space/space-time. Secondly, eliminativism seems to block natural ways to categorize many entities introduced in scientific theories as relations or relation-like beings.

Considering first the metaphysics of space/space-time, spatio-temporal relations are widely considered as external relations between objects. In other words, their holding is contingent relative to the existence of their relata. Since eliminativists deny the existence of relations, they would be obliged to consider (contingent) spatio-temporal relations *derived internal relations*, internal relations that hold due to the holding of the internal relations between entities some of which are distinct from the relata of the original relation (Keinänen, Keskinen & Hakkarainen 2017, sec. 2). For instance, having the same mass as between objects *a* and *b* is a derived internal relation, which holds because of *a* and *b* having equal ("exactly similar") mass tropes as their certain kinds of parts. Similarly, the spatial distance between objects *a* and *b* might be contingent relative to the existence of *a* and *b* if there are certain additional, mutually internally related entities also internally related to both *a* and *b*.

Most of the recent eliminativist views about relations have been committed to a *substantivalist theory of space/space-time* (a *substantivalist view* for short): the claim that space-time points, regions of space-time or space-time itself are primitive object-like entities.[17] The general idea of these eliminativist accounts

bestow them with any additional category features. Therefore, Simons' earlier account of relational tropes is seriously incomplete (cf. note 10).

[16] Certain advocates of primitive substances, such as Heil (2012, 2016) and Lowe (2016), have also proposed strategies to avoid the postulation of relations.

[17] Lowe's (2016) otherwise interesting account is a case in point. Mulligan (1996) avoids commitment to ("thick") spatio-temporal relations by as-

is that the spatio-temporally related entities stand in some internal relations (such as identity or monadic inherence) to (the parts of) the space-time structure. The contingency of spatio-temporal relations is either explained by means of the contingent existence of the space-time structure or simply denied. Since trope theories strive to eliminate objects (in the sense of bearers of properties or relations) as a primitive category, a trope theorist adopting a substantivalist view would be obliged to construct space/space-time by means of tropes. No clear idea of such construction has been presented so far. Substantivalist theories of space/space-time typically allow for the existence of empty space-time points. No substantivalist trope theorist has managed to show that empty space-time points can be constructed by means of tropes.

Thus, an *anti-substantivalist* or a *relationalist theory of space/space-time* might seem to be a preferable view for the trope theorist.[18] Peter Simons (2016) has proposed a construction of space-time by means of internal relations among the fundamental concrete entities. This view is better characterized as an anti-substantivalist than a relationalist account of space-time because it does not introduce any relations or other relational entities. Instead, Simons assumes that fundamental entities are *occurrents* (i.e., processes and events) having their spatio-temporal locations necessarily. All standard continuants (or, endurants) are *Fregean abstractions* from occurrents.[19] Moreover, he assumes a causal theory of time.

Simons' general claim that all fundamental particulars are occurrents is contestable and his examples of the construction

suming space-time points, which have tropes as their individual accidents. Although being a trope theorist, Campbell is also inclined to adopt a substantivalist theory of space-time. In his Abstract Particulars (1990), Campbell rejects his earlier (Campbell 1981) identification of tropes with "formed volumes", i.e., parts or regions of space/space-time. In his final, scientifically inspired version of trope theory, Campbell takes space-time as a single simple entity, and all other entities are fields in the same space-time manifold (1990, 145ff.).

[18] All theories of space/space-time that deny the existence space/space-time or its parts as a separate substance(s) are anti-substantivalist.

[19] Cf. Simons (2000, 2008) for a proposal to construct continuant objects as Fregean abstractions from occurrents.

of continuants from occurrents have remained schematic.[20] Even if all fundamental entities were occurrents, we would need additional reasons to support the claim that occurrents have their specific locations necessarily. It might be tempting to *individuate* processes and events by their spatio-temporal location, but it is not clear whether such intuitions about individuation are applicable to a process ontology like the one suggested by Simons.[21] It seems to be a safer alternative for a trope theorist to adopt a relationalist theory of space/spacetime, which takes (some of the) spatio-temporal relations as relational entities. One can reconcile this full-blown relationalist theory of space/space-time with a more standard view that the same entities could have had different locations or relative positions. In other words, the spatial/spatio-temporal relations between entities are contingent relative to the existence of their relata.

Finally, the current science and the current quantum physics in particular provide trope theorists independent reasons to postulate relations or relation-like entities. The current quantum physics introduces entangled states of two- or multi-particle systems, which are serious candidates for fundamental relations between particles (cf. Teller 1986; Karakostas 2009). For instance, Paul Teller (1986, sec.4) has argued that entangled spin-states of two superposed electrons are best considered as relations, which do not supervene on the spatio-temporal arrangement and the monadic properties of these particles. In the context of trope theory, these entangled spin-states would be good candidates for relational tropes (cf. Keinänen 2011, 434). Additionally, a trope theorist may need to introduce relational tropes to account for the "emergent" features of complex objects, that is, the features of complex objects which do not supervene on the properties of their proper parts.[22] Finally, the present-day quantum physics in-

[20] For instance, Simons' (2000) examples of constructed continuants are complex objects. Nevertheless, it remains unclear whether we would need continuant objects not reducible to occurrents as proper parts of complex objects.

[21] Cf. MacBride (2016, ch.2) for a brief criticism of Simons' eliminativism.

[22] One possible example of such emergent properties are masses of complex physical particles like helium atoms, which cannot be directly reduced to the masses of their proper parts. I have suggested elsewhere that

troduces virtual particles (such as photons and gluons) to account for interactions between micro-particles (electrons, quarks). It is an interesting, a hitherto unstudied option to consider such interactions relational tropes.

Even this limited set of examples shows that there is reasonable work for relational tropes in an a posteriori orientated trope theory. Most importantly, relational tropes (or tropes which would function like relational tropes) would bestow trope theory with the required ontological explanatory power to respond to the challenge of the different, currently popular relational ontologies. Given the serious difficulties the relata-specific relational tropes face, the trope theorist is advised to seek for a reductive analysis of relational inherence.

4. The analysis of relational inherence

The basic idea in the reductive analysis of relational inherence is to generalize the trope theoretical analysis of monadic inherence to "relational tropes". In the analysis of relational inherence, the general goal is to provide a metaphysical reduction of relational inherence: to identify the facts about two or more entities being connected by a relation with the facts about the entities of the trope theoretical category system. Since relational inherence is explained away, also relational tropes (i.e. primitively relating entities) are eliminated from trope theory. However, certain tropes, which I call "*r-tropes*", take the role of relational entities in the present account. The main difference between standard "property tropes" and r-tropes is their standing in slightly different kinds of formal ontological relations and being parts of different kinds of complex entities. Nevertheless, there is no such thing as a primitive category distinction between primitively characterizing entities (properties), on the one hand, and primitively relating entities (relations), on the other.[23]

such emergent properties are best categorized as relational tropes, cf. Keinänen (2011, 447).

[23] The entities belonging to the same category bear the same formal ontological relations to themselves and to certain other entities. These formal ontological relations are internal relations – necessarily, if certain entities exist, it is a primitive fact about them that certain formal ontological rela-

Hence, the reductive analysis will have two main goals: the first is to eliminate the *primitive* distinction between relational tropes and their relata, which threatens to set serious limitations to the ontological explanatory power of trope theories. The second goal is to incorporate the relation-like entities, which are capable of serving the core functions set to relations in an a posteriori basis, into the trope theoretical framework. My goal will not be to deal with all conceivable cases of relations. In order for the reductive analysis of relational inherence to serve its purpose, it suffices to consider credible a posteriori examples of relational entities and submit their relational inherence to reductive analysis.

Recall that the different trope bundle theories analyze *monadic inherence* in different ways. For the present purposes, it suffices to consider two trope theories. Campbell's (1990) theory takes objects as mereological sums of mutually co-located ("compresent") tropes. Correspondingly, trope *t* inheres in object *i* if and only if *t* is a part of *i* and *t* is co-located with *i*.[24]

By contrast, in the trope theory SNT (Keinänen 2011; Keinänen & Hakkarainen 2010, 2014), tropes are assumed to be mutually existentially dependent beings and objects are constituted as aggregates of tropes connected by the formal ontological relations of rigid and generic dependence.[25] Here, I confine myself to outlining the features of the SNT directly relevant to the present discussion.[26] According to the SNT, every object has either a single nuclear trope or, alternatively, two or more tropes rigidly dependent on each other, the *nu-*

tions hold, cf. Hakkarainen (2018) for this kind of account of ontological categories.

[24] Since Campbell (1990, secs. 4.3-4.4) constructs complex quantity tropes as "conjunctive compresences" of simpler tropes falling under the same determinable, an additional maximality condition would be needed to be added to the analysis in order to deal with such mutually co-located tropes forming a complex trope.

[25] Let "$\leq$ " be a relation of improper parthood and "E!" the predicate of (singular) existence. Entity *e* is strongly rigidly dependent on entity *f*, if the following condition holds: $\neg(\Box\, E!f) \,\&\, \Box\, ((E!e \rightarrow E!f) \,\&\, \neg(\, f \leq e\,))$, cf. Simons (1987, 112, 294ff.).

[26] Cf. Keinänen (2011, sec. 4) for a more systematic presentation of the SNT.

clear tropes.[27] Nuclear tropes are necessary parts of an object *i* and, intuitively, constitute its "necessary properties". Trope *t* is a part of object *i* if and only if *t* is rigidly dependent only on the nuclear tropes of *i*. Object *i* is a *dependence closure* of tropes with respect to rigid dependence.[28] Because object *i* is a dependence closure of tropes, *i* is not rigidly dependent on any entity which not its proper part.[29]

Unlike Campbell's trope theory, the SNT does not build objects by means of co-location ("compresence") but uses the relations of existential dependence.[30] The second major difference between these two trope theories concerns the determination of the location of individual tropes. In Campbell's trope theory, individual tropes are relata of the basic spatio-temporal relations, whereas in the SNT, this function is given to certain trope bundles. According to the SNT, the certain kinds of aggregates of tropes (e.g. the nuclear tropes of a substance) form individuals, which are minimal relata of the basic spatio-temporal relations. The spatio-temporal locations of these complex entities determine the locations of their constituent tropes. In a simple case, object *i* is constituted solely by its nuclear tropes and the location of *i* determines the location of the tropes that are its proper parts. The SNT analyzes monadic inherence in this special case as follows: trope *t* is a

[27] According to the SNT, trope *t* is a nuclear trope if and only if 1) *t* is not rigidly dependent on any other trope (a single nuclear trope), or 2) *t* is rigidly dependent on certain trope(s) which are also rigidly dependent on *t* (two or more nuclear tropes).

[28] A dependence closure of tropes with respect to rigid dependence is a plurality of tropes in which all rigid dependencies of the tropes in the plurality are fulfilled. Moreover, we assume that necessarily, if these tropes exist, they form an individual. As a consequence, that individual is not rigidly dependent on any mereologically disjoint entity, cf. Keinänen (2011, 446-447).

[29] The applicability of the notion of rigid dependence is restricted to contingent existents cf. note 25. Moreover, as advocates of naturalistic nominalism (cf. section 3), trope theorists can reject the existence of sets, on which objects would (allegedly) be rigidly dependent.

[30] Here, I offer only a simplified sketch of the construction of objects in the SNT.

property of object *i* if and only if, necessarily, if *t* exists, *t* is a proper part of *i* and *t* is co-located with *i* [31]

In what follows, my strategy is to generalize the analysis of monadic inherence of the trope theory SNT to *r-tropes*. Moreover, the analysis will adopt one of the main assumptions of the relata specific answer, namely, that r-tropes are multiply rigidly dependent (MRD, for short) on two or more entities. Since multiple rigid dependence is not sufficient to relational inherence, we need to specify additional conditions that hold of trope *r* and objects *a* and *b* if *r* relationally inheres in *a* and *b*. Rigid dependence will be supplemented by the condition of necessary co-location as we will see below. Although the analysis of relational inherence is based on the idea that r-tropes are dependent existents, it is purported to be consistent with considering tropes other than r-tropes existentially independent as Williams (1953) and Campbell (1990) do.

Thus, r-tropes are *multiply rigidly dependent* (MRD) on two or more entities. Assume, for instance, that r-trope *r* is a 1 m distance trope connecting entities *a* and *b*: *a* and *b* are in a 1 m distance from each other. Trope *r* is MRD on *a* and *b*. This multiple rigid dependence involves three things. First, necessarily, if distance trope *r* exists, entities *a* and *b* (its "relata") also exist. Second, entities *a* and *b* are mereologically disjoint and mereologically disjoint from *r*. In other words, r-tropes connect mutually "wholly distinct" (mereologically disjoint) entities and are wholly distinct from the entities which they connect, their "relata". Third, entities *a* and *b* are not rigidly dependent on each other. The third condition rules out the cases in which trope *r* is rigidly dependent on the nuclear tropes of a single object. Finally, in order to rule out trivial cases (e.g., in which the dependees *a* or *b* are necessary existents), it is presupposed in the characterization of MRD that trope *r* and entities *a* and *b* are all contingent existents.[32]

[31] Keinänen (2011, 438-440). The more general condition, which also deals with the tropes contingent to an object, is temporally qualified: necessarily, trope *t* is co-located with *i* when it exists (ibid. 440ff.).

[32] The characterization of rigid dependence and multiple rigid dependence are thus restricted to contingent existents, cf. Simons (1987, 294ff.) for a similar restriction.

The crucial step in the analysis is to add three more conditions in order to obtain the conclusion that trope *r* relates, that is, relationally inheres in *a* and *b*. The first two conditions concern the constitution of an r-complex. The first is that *a* and *b* are the only entities on which trope *r* is rigidly dependent, *r* is rigidly dependent only on *a* and *b*. Secondly, trope *r* together with its dependees ("relata"), *a* and *b*, form an individual, which I call an "*r-complex rab*".[33] R-complex *rab* is a dependence closure of its proper parts with respect to rigid dependence. As a dependence closure of its parts, r-complex *rab* is itself a strongly rigidly independent entity, it is not rigidly dependent on any entity that is mereologically disjoint from *rab*. Hence, r-complexes are substances in the weak sense of being *strongly independent particulars and individuals*.

The third condition is that r-complex *rab* is a spatio-temporally located entity: r-complex *rab* has a spatio-temporal location and its location determines the location of its constituent r-trope, 1 m distance trope *r*. Like the objects constituted by their nuclear tropes, an r-complex is a strongly independent particular and has all of its proper parts necessarily. Moreover, as in the case of objects having only nuclear tropes, the location of the r-complex determines the location of its existentially dependent part, r-trope *r*. As we will see below, some, but not all, r-complexes are entities that figure in the basic spatio-temporal relations and have an independent location in this sense. Again, they are like objects constituted by nuclear tropes. On the basis of these assumptions, I now propose the following analysis of the holding of relational inherence:

[RI]:

Trope *r* relationally inheres in *a* and *b* if and only if:

1. *r* is multiply rigidly dependent (MRD) on *a* and *b*, but not rigidly dependent on any entity that is not a part of *a* or a part of *b*.
2. *a* and *b* are not rigidly dependent on *r*.
3. *a* is not rigidly dependent on *b*, and *b* is not rigidly dependent on *a*.

[33] Note that every r-complex is an individual and a mereological sum of its parts (e.g., r + a + b = s).

4. *r*, *a* and *b* constitute an individual, r-complex *rab*.
5. Necessarily, if *r* exists, *r* is exactly co-located with *rab*.

Let us take again 1 m distance trope *r* as an example. Trope *r* relates (relationally inheres in) *a* and *b*, if *r* is both multiply rigidly dependent on *a* and *b* and necessarily (exactly) co-located with r-complex *rab*, which is a mereological sum of all these three entities (i.e., r+a+b).[34]

The purpose of [RI] is to generalize the analysis of monadic inherence of the trope theory SNT to r-tropes, that is, the tropes that fulfil clauses 1-3 of [RI]. This generalization is achieved by assuming that the corresponding r-complex, whose existence is entailed by the existence of *r*, is an individual having a specific spatio-temporal location. Moreover, like the location of an individual constituted by mutually rigidly dependent tropes (nuclear tropes), the location of the r-complex determines the location of its existentially dependent parts (an r-trope in this special case). Thus, necessarily, if r-trope *r* exists, it is co-located with *rab*. As a consequence, trope *r* fulfills the conditions of monadic inherence in relation to complex *rab*: necessarily, if *r* exists, *r* is a (proper) part of *rab* and *r* is co-located with *rab*. Thus, *r* is a monadic property of complex *rab*. According to [RI], by being a monadic property of r-complex *rab*, trope *r* also relationally inheres in *a* and *b*.

In order to motivate this analysis of relational inherence, it is useful to begin with the idea of tropes as particular natures (-e charges, 1 m lengths, etc.). According to the analyses of monadic inherence discussed above, tropes are monadic properties of an individual because they are mutually co-located parts of that individual, which might also need to fulfil some additional conditions (as in the SNT). R-tropes, like 1 m distance trope *r*, are particular natures co-located with the corresponding r-complexes and monadic properties of these r-complexes. Furthermore, r-trope *r* is a certain kind of entity that connects mutually distinct entities, *a* and *b*, into a certain kind of more inclusive whole. In order to see this, we need to observe three things. First, trope *r* and complex *rab* are (weak-

[34] In what follows, I leave out the qualification, although I refer to exact co-location when talking about "co-location".

ly) rigidly dependent on *a* and *b*. Thus, second, given that trope *r* exists, *a* and *b* are parts of a certain kind of r-complex, *rab*. Third, since *a* and *b* are proper parts of complex *rab,* their locations are parts of the location of *rab*.[35] Consequently, locations of *a* and *b* are parts of the location of trope *r*.

Hence, according to [RI], tropes relate (relationally inhere in) their relata by being properties of the respective r-complexes ("relational complexes"), which have their relata as proper parts. In the special case discussed just above, trope *r* (1 m distance trope *r*) relates entities *a* and *b* in a certain way because *r* "makes" *a* and *b* as parts of a certain kind of complex individual, 1 m distance r-complex *rab*.

5. Asymmetric and non-symmetric relations

An obvious worry with [RI] concerns *asymmetric* and *non-symmetric* relations. Causal relations, relations of spatial direction (such as being to the left of) and temporal direction (being after than) are salient examples of asymmetric relations. Many relations between quantitative properties (such as being greater than or equal to), some spatial relations (facing) and relations manifesting human attitudes (admiring, loving) are *non-symmetric* without being asymmetric. Prima facie, asymmetric and non-symmetric relations hold between entities in a certain order (cf. Fine 2000, 1). For instance, Muodoslompolo is to north of Tornio but Tornio is not to north of Muodoslompolo (asymmetry); Young Werther loves Charlotte, but Charlotte does not love Werther (non-symmetry). It seems that [RI] is not able to deal with non-symmetric or asymmetric relations because r-tropes do not themselves bestow any order on the parts of r-complexes. Therefore, it seems that clause [RI] can only provide us with an account of the special cases of relational inherence in which the relation under consideration is *symmetric* (e.g., the relational tropes of spatial distance if there are such entities). As a consequence, if we accept the proposed analysis, we seem to be obliged to deny the existence of all asymmetric and non-symmetric relations. This is an untenable conclusion

[35] As Parsons (2007, 213) argues, all concrete entities satisfy the following principle of Expansivity: the spatial location of a whole is as least as inclusive as the spatial location of its proper parts.

if we take seriously the examples of relations the empirical science gives us (cf. MacBride 2014, sec.1, 2016, sec. 4).

Nevertheless, some of the above examples are *basic* or *derived internal relations*, which do not exist as separate relational entities. Rather, tropes and complex entities they constitute are internally related in different ways.[36] In section 3, I already mentioned derived internal relations. Having a greater mass than or having a smaller charge than are examples of asymmetric *derived internal relations*, which hold between objects having certain kinds of mass or charge tropes as their parts. Moreover, the quantity tropes falling under a determinable (e.g. electric charge) are mutually connected by the different basic internal relations of proportion (e.g., 1:1 proportion or -3:1 proportion) and the basic internal relation of order (greater than or equal to). These basic internal relations hold because tropes are certain thin particular natures - the existence of the related entities is a sufficient condition for their obtaining. Moreover, the holding of these relations does not depend, even indirectly, on the existence of any specific entities distinct from their original relata (Keinänen; Keskinen & Hakkarainen 2017, sec.3). Here, the relation of order is non-symmetric, whereas the relations of proportion are symmetric or asymmetric.

Formal ontological relations constitute additional examples of basic or derived internal relations.[37] For instance, tropes are proper parts of objects, which is an asymmetric formal ontological relation. Moreover, in the SNT, all tropes constituting an object are connected by the non-symmetric formal ontological relation of rigid dependence. Asymmetric and non-symmetric basic or derived internal relations do not cause any problem for the present analysis: because they are not relational entities, internal relations do not relationally inhere in anything. Rather, it is a primitive fact about quantity tropes that they are ordered, that e charge tropes are great-

[36] Cf. Keinänen; Keskinen & Hakkarainen (2017, sec.2) for a more precise characterization of the distinction between basic and derived internal relations.

[37] Cf. Hakkarainen & Keinänen (2017) for the distinction between formal ontological relations, which are "nature neutral", and other basic internal relations.

er than e/3 charge tropes, for instance. Similarly, it is a primitive fact about tropes that they are rigidly dependent on certain distinct tropes.

Moreover, I adopt a sparse theory of relational entities, which is in line with a sparse theory of tropes (Campbell 1990, sec. 1.8): there are only few different kinds of relational entities, which are all discovered empirically. An advocate of a sparse theory of relational entities can remain skeptical of the existence of any such macro-level relational entities as the relational tropes of love or macroscopic causation (Simons 2003; Lowe 2016, 106-110). The best prima facie candidates for r-tropes are basic (or, comparatively basic) physical quantities. Among them, there are asymmetric vector quantities like momentum and asymmetric quantitative spatial and temporal relations.

Assuming that all r-tropes are quantities, we can present a general strategy to deal with their asymmetry. In this account, we need not assume that inherence of r-tropes is asymmetric. In order to take a simple example, consider distances in some direction in a one-dimensional space.[38] Assuming that there are distance-direction tropes, they are vector quantities, magnitudes with a certain direction. In predicate logic, the direction of an asymmetric relation is typically indicated by argument places. Thus, for instance, object *a* is 1 m to the left of object *b*, Lab. Sentence "Lba" can be used to indicate that *b* is 1 m to the left of *a*. Hence, a relational predicate applies to a pair of objects in different ways depending on the direction of the corresponding relation.

It is important to keep in mind that r-tropes do not have any formal-ontologically specified direction. First, r-tropes do not have any argument places, by means of which the relata are put into some order. Second, the source of the order of the relata cannot be the different ways in which an r-trope is multiply rigidly dependent on certain entities. There is only one and a unique way in which an r-trope is multiply rigidly dependent on certain entities.

[38] Of course, space-time intervals have replaced distances as basic quantities in the current physical theories of space-time. Therefore, I present this example of distance direction tropes only as an illustration.

Nevertheless, the r-tropes of distance-direction are, as particular natures, determinate magnitude-directions (vectors). Like all quantity tropes falling under a determinable, the r-tropes of distance-direction are mutually connected by the different basic internal relations of positive or negative proportion (like, say 1:1 proportion or -3:1 proportion) and the basic internal relation of order (greater than or equal to). The choice of the unit for distance-direction is a matter of convention as well as which of these r-tropes of distance-direction get positive and which negative values. By contrast, because of being determined by the distance-direction tropes, the relations of proportion and order between distance-direction tropes remain invariant in all choices of the unit.[39]

Whether two r-tropes of distance-direction are connected by a relation of positive or negative proportion spells out their relative directions. The r-tropes connected by some relation of negative proportion are distance-directions to opposite directions, whereas the distance-direction r-tropes to the same direction are connected by some relation of positive proportion. Thus, according to the present approach, the direction is already included in a distance-direction trope as a particular nature. Similarly, an r-complex having a distance-direction trope as a proper part has an intrinsic direction determined by the respective r-trope, which may be opposite to the direction of another r-complex.

Hence, the present approach denies that r-tropes have any formal-ontologically determined (absolute or relative) direction. Unlike the recent views in the metaphysics of relations (e.g., positionalism or anti-positionalism), the present approach does not introduce any general (logical or formal-ontological) devices to determine the relative direction of argument places (cf. Fine 2000, secs. 3-4; MacBride 2014). Instead, the direction of a relational fact is determined by an r-trope as a particular nature.[40] The present approach does not

[39] Cf. Keinänen; Keskinen & Hakkarainen (2017, sec.3) for a defense of the same general account of internal relations between quantity tropes falling under a determinable.

[40] Certain r-tropes have an absolute direction as vectors. However, the direction of an r-trope is based on its nature and it does not correspond to any fixed order of the relata figuring as arguments of a relation. Similarly,

over-generate directionality because the non-directional r-tropes falling under a determinable (e.g., distance tropes if there are such entities) are related only by the relation of order and the relations of positive proportion.

The above kind of quantitative r-tropes are good prima facie candidates for truthmakers of asymmetric predications such as "a is 1 m to the left of b" or "b is 1 m to the right of a". According to the present conception, these two sentences have the same truthmaker (i.e., some r-complex *rab*), but they correspond to the different ways in which the positive/negative unit of distance-direction can be selected.

Nevertheless, the best current candidates for the basic spatio-temporal r-tropes are particular space-time intervals. They are mutually connected by the different relations of positive, negative or zero proportion. However, space-time intervals do not have any intrinsic direction. Rather, the different kinds of intervals between objects in space-time points indicate, for instance, whether or not these space-time points can be involved in one temporally continuous succession of events. Thus, we are entitled to expect that asymmetric predications like "a is before b" or "a causes b" do not have r-tropes as their sole truthmakers, but, rather, more complicated structures of entities, which may involve some r-tropes.[41]

6. The location of r-tropes

According to clause [RI], an r-complex is an individual possessing certain spatial or spatio-temporal location, which determines the location of the corresponding r-trope. An advocate of the present analysis of relational inherence is obliged to provide some account of the determination of the location of r-complexes. Providing an answer to this question is particularly important in the case of r-complexes partially constituted by spatial or spatio-temporal r-tropes. There is a threat of a regress of spatial or spatio-temporal r-tropes if we

there is no fixed way to indicate this direction by means of the order of the argument places of a two-place predicate, for instance.

[41] For instance, the claims about temporal precedence of events might be made true by complicated physical facts involving the increase of total entropy in universe.

need to postulate additional r-tropes to account for the location of every such r-complex.

The second issue concerns the peculiar character of spatial r-tropes. As we saw above, spatial r-tropes are assumed to be distances or distance-directions between the different occupants of space. We need no recourse to relational inherence in the formal-ontological characterization of the r-complexes partially constituted by the spatial r-tropes. Nevertheless, since spatial r-tropes are assumed to be distances *between* objects, one might claim that being a relation is smuggled into the (non-formal) nature of r-tropes.[42] In what follows, I deal with these two issues concerning the spatial or spatio-temporal r-tropes and the respective r-complexes first. Finally, I address the determination of the location of r-complexes constituted by means of other kinds of r-tropes.

The current metaphysical discussion of space-time is, in large part, still dominated by the rivalry between substantivalist and relationalist theories about space-time.[43] According to the contemporary substantivalists, space-time is an independently existing entity of its own, which is constituted by space-time points having certain inertial features like curvature (Teller 1991, 363-4, 379). Relationalism (or, "liberalized relationalism" as Teller calls it) introduces spatio-temporal relations between actual objects and actual and possible objects. One is supposed to obtain the empty space-time points as locations of possible objects. Moreover, one is supposed to be able to construct the whole space-time manifold (the system of space-time points) by means of spatio-temporal relations (*ibid*).

From the point of view of trope theory, both of substantivalism and relationalism about space-time are problematic views. In section 3, I already mentioned the difficulty of constructing empty space-time points by means of tropes. A related problem can be addressed to liberalized relationalism: it is reasonable to demand that relations can connect only entities that exist. Thus, relationalism is prima facie committed to the existence of possible but non-actual objects. The merely possible objects are needed as relata of

[42] I am grateful to Jani Hakkarainen for presenting this problem.

[43] For additional alternative accounts of space-time, cf. e.g., Pooley (2005).

spatio-temporal relations. It is difficult to present any account of the construction of merely possible objects from tropes, which are actual and spatio-temporal entities. As a consequence, liberalized relationalism seems to be an equally unacceptable conception of space-time for a trope theorist as substantivalism, with which it is supposed to compete.

Without solving the problem of empty space-time points here, I adopt a broadly relationalist conception of space-time. According to it, r-tropes, which correspond to spatio-temporal relations, and the respective r-complexes *constitute* space-time (space might be used in illustrations). In other words, space-time is not considered as a separate object. Rather, space-time is a structure (wholly or partially) constituted by the mutually connected r-complexes. Although there are open issues in this type of view (like the status of empty space-time points if there are such items), it seems to provide us with a promising starting point for the construction of space-time from tropes.

For purposes of illustration, let us consider space and spatial relations between objects (distances or distance-directions). Consider now a single r-complex *rab*, which is a part of space, that is, the r-complex which has trope *r* (certain particular distance or distance-direction), and objects *a* and *b* as its parts. We can identify r-trope *r* with the shortest path of space connecting *a* and *b*. Trope *r* is a particular nature, a certain length in space. By being rigidly dependent on *a* and *b* and co-located with the respective r-complex, trope *r* can exist only in presence of the contents of space (space-time).

The location of r-complex *rab* is determined holistically, by its place in the system of spatial r-complexes. Assume that all other r-complexes than *rab* exist, among them the r-complexes that overlap *rab* by having *a* or *b* as their parts. If these other r-complexes exist, there must also be an r-complex connecting *a* and *b*. In other words, there must be an r-complex which has the same position in the network of r-complexes as *rab*. If *rab* exists, it has this specific position. Thus, the system of r-complexes determines the location of *rab* as a part of space.

It is possible to make additional assumptions, which constrain the nature of r-complex *rab* or any other r-complex having the same place in the system of r-complexes. In a special case of Euclidean space, the other existing r-complexes are parts of space and the spatial relations between *a* and other objects, and *b* and other objects are sufficient to necessitate the fact that *a* and *b* are connected by a r-trope (particular distance or distance-direction) of a certain determinate kind. However, the structure of space may have local variation, which allows for *a* and *b* to be connected by different kinds of r-tropes. The identification of relational tropes with paths of space (space-time) solves the location problem of spatial (spatio-temporal) r-tropes: they are concrete entities that contribute to constituting space (space-time).

The second problem concerns the alleged primitive relatedness included in the nature of a spatial r-trope. In response, one can avoid primitive relatedness in the following way: objects *a* and *b* are parts of r-complex *rab*. Because of being proper parts of the distinct r-complexes, the locations of these objects, *a* and *b*, are proper parts of the locations of the distinct r-complexes. The r-complexes, which have an object as a proper part, assign to the object a determinate location as an intersection of the locations of these r-complexes. Therefore, we need not assume that an r-complex determines a primitive between-ness relation connecting objects *a* and *b*; rather, the system of r-complexes determines that objects *a* and *b* are in a certain distance (distance-direction) from each other.[44]

In the end of section 3, I provided some prima facie examples of relational entities such as entangled spin-states of multi-particle systems, emergent properties of complex objects and virtual particles. They are both good candidates for r-tropes and spatially located entities. It seems that the respective r-complexes are independently located entities and that their locations can be determined by spatial/spatio-temporal r-tropes. Of course, the specific details of such an account must be worked out in distinct cases.

[44] One might claim that r-complexes self-locate (are their own locations). However, this not quite right because we need the whole system of r-complexes for an r-complex to have a specific location.

7. Conclusion

Because of the reductive analysis of monadic inherence (objects having tropes), trope theories have promised to analyze away the primitive dichotomy between characterizing (properties) and characterized entities (objects). As I argued in section 2, the best trope theoretical account of relations, the relata specific view, re-introduces the same dichotomy at the level of relations. This is unsatisfactory and it reduces the initial appeal of trope theories. Nevertheless, we need relation-like entities in an adequate conception of the categorial structure of reality, which rules out eliminativism about relations (section 3).

Therefore, in section 4, I presented a novel trope theoretical analysis of relational inherence, which is a generalization of the analysis of monadic inherence provided by the trope theory SNT. The analysis provides us with a metaphysical reduction of relational inherence to the facts about the entities of the trope theoretical category system. The core feature of the analysis is to introduce multiply rigidly dependent tropes, which I call *r-tropes*. Like all tropes, r-tropes are particular natures with a specific location. If r-trope *r* is multiply rigidly dependent on objects *a* and *b*, entities *r*, *a* and *b* form a complex individual, r-complex *rab*. An r-complex is a concrete particular and the location of r-complex *rab* determines the location of *r*. R-trope *r* relationally inheres in entities *a* and *b* by unifying them into r-complex *rab* and by being colocated with *rab*. For instance, since 1 m distance trope *r* unifies objects *a* and *b* into complex *rab*, objects *a* and *b* are in 1 m distance from each other.

In section 5, I argued that the present analysis can deal with asymmetric and non-symmetric relations by assuming that all fundamental relations are quantities. Finally, section 6 delivers an account of the determination of the location of r-tropes also in the difficult case in which an r-trope contributes to determining the spatial or spatio-temporal location of objects.[45]

[45] I wish to thank A.R.J. Fisher, Robert Garcia, Daniel Giberman, Jani Hakkarainen, Antti Keskinen, Anna-Sofia Maurin, Tommi Vehkavaara and two anonymous reviewers for helpful comments on the earlier drafts of this article. Moreover, I thank the participants of the Nordic Network of

University of Tampere

References

Allen, S. (2016), *A Critical Introduction to Properties*, London, Bloomsbury.

Armstrong, D.M. (1989), *Universals – an Opinionated Introduction*, Boulder, Westview Press.

Betti, A. (2015), *Against Facts*, Cambridge Ma., MIT Press.

Campbell, K. K. (1981), "The Metaphysic of Abstract Particulars", *Midwest Studies in Philosophy* 6, pp. 477–488.

Campbell, K. K. (1990), *Abstract Particulars*, Oxford, Basil Blackwell.

Ellis, B. D. (2001), *Scientific Essentialism*, Cambridge, Cambridge University Press.

Fine, K. (2000), "Neutral Relations", *The Philosophical Review* 109(1), 1-33.

Fisher, A. R. J. (2018), "Instantiation in Trope Theory", *American Philosophical Quarterly* 55(2), pp. 153–164.

Giberman, D. (2014), "Tropes in Space", *Philosophical Studies* 167(2), pp. 453–472.

Hakkarainen, J. (2018), "Ontological Form and Moderate Categorial Realism", manuscript.

Hakkarainen, J. & Keinänen. M. (2017), "The Ontological Form of Tropes – Refuting Douglas Ehring's Main Argument Against Standard Trope Nominalism", *Philosophia* 45(2), pp. 647–658.

Heil, J. (2012), *The Universe As We Find It*, Oxford, Oxford University Press.

Heil J (2016), "Causal Relations", in A. Marmodoro & D. Yates (eds.), *The Metaphysics of Relations*, Oxford, Oxford University Press, pp. 127–137.

Karakostas, V. (2009), "Humean Supervenience in the Light of Contemporary Science", *Metaphysica* 10(1), pp. 1–26.

Keinänen, M. (2011), "Tropes – the Basic Constituents of Powerful Particulars?", *Dialectica* 65(3), pp. 419–450.

Keinänen, M. (2018), "Instantiation and Characterization: Problems in Lowe's Four-Category Ontology", in Timothy Tambassi (ed.), *Studies in the Ontology of E.J. Lowe*, Neunkirchen-Seelscheid, Editiones Scholasticae, pp. 109–124.

Keinänen, M. & Hakkarainen, J. (2010), "Persistence of Simple Substances", *Metaphysica* 11(2), pp. 119–135.

Metaphysics conference "Metaphysics of Properties". Finally, I thank Finnish Cultural Foundation for funding this research.

Keinänen, M. & Hakkarainen, J. (2014), “The Problem of Trope Individuation – A Reply to Lowe”, *Erkenntnis* 79(1), pp. 65–79.

Keinänen, M., Keskinen, A. & Hakkarainen, J. (2017), “Quantity Tropes and Internal Relations”, *Erkenntnis*, published online.

Lowe, E. J. (2006), *The Four-Category Ontology*, Oxford, Oxford University Press.

Lowe, E. J. (2009), *More Kinds of Being*, Oxford, Wiley-Blackwell.

Lowe, E. J. (2015), “In Defence of Substantial Universals”, in G. Galluzzo & M. J. Loux (eds.), *The Problem of Universals in Contemporary Philosophy*, Cambridge, Cambridge University Press, pp. 65–84.

Lowe, E. J. (2016), “There are probably no relations”, in A. Marmodoro & D. Yates (eds.), *The Metaphysics of Relations*, Oxford, Oxford University Press, pp. 100–112.

MacBride, F. (2011), “Relations and Truth-Making”, *Proceedings of the Aristotelian Society* 111, pp. 159–76.

MacBride, F. (2014), “How Involved You Want To Be In A Non-Symmetric Relationship”, *Australasian Journal of Philosophy* 92, pp. 1–16.

MacBride, F. (2016), "Relations", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/win2016/entries/relations/>.

Martin, C. B. (1980), “Substance Substantiated”, *Australasian Journal of Philosophy* 58(1), pp. 3–10.

Maurin, A-S. (2002), *If Tropes*, Dordrecht: Kluwer Academic Publishers.

Maurin, A-S. (2010), “Trope Theory and the Bradley’s Regress”, *Synthese* 175(3), pp. 311–326.

Maurin, A-S. (2011), “An Argument for the Existence of Tropes”, *Erkenntnis* 74(1), pp. 69–79.

Mulligan, K. (1998), “Relations Through Thick and Thin”, *Erkenntnis* 48 (2 & 3), pp. 325–353.

Parsons, J. (2007), “Theories of Location”, in D. W. Zimmerman (ed.), *Oxford Studies in Metaphysics*, vol. 3, pp. 201–232.

Pooley, O. (2005), “Points, particles, and structural realism”, in Rickles, D., French, S. & Saatsi, J. (eds.), *The Structural Foundations of Quantum Gravity*, Oxford, Oxford University Press, pp. 83–120.

Simons, P. M. (1987), *Parts – a Study in Ontology*, Oxford, Clarendon Press.

Simons, P. M. (2000), “Continuants and Occurrents”, *Proceedings of the Aristotelian Society, Supplementary Volume* 74, pp. 59–75.

Simons, P. M. (2003), “Tropes, Relational”, *Conceptus* 35, pp. 53–73.

Simons, P. M. (2008), “The Thread of Persistence”, in Kanzian, C. (ed.), *Persistence*, Frankfurt, Ontos Verlag, pp. 165–184.

Simons, P. M. (2010), "Relations and Truth-Making", *Proceedings of the Aristotelian Society, Supplementary Volumes 84,* pp. 199–213.

Simons, P. M. (2014), "Relations and Idealism: On Some Arguments of Hochberg against Trope Nominalism", *Dialectica*, 68, pp. 305–315.

Simons, P. M. (2016), "External Relations, Causal Coincidence, and Contingency", in A. Marmodoro & D. Yates (eds.), *The Metaphysics of Relations*, Oxford, Oxford University Press, pp. 113–126.

Teller, P. (1986), "Relational Holism and Quantum Mechanics", *British Journal for the Philosophy of Science* 43, pp. 201–218.

Teller, P. (1991), "Substance, Relations, and Arguments about the Nature of Space-Time", *Philosophical Review* 100(3), pp. 363–397.

Wieland, J. W. & Betti, A. (2008), "Relata-specific Relations: A Response to Vallicella", *Dialectica* 62(4), pp. 509–524.

Williams, D. C. (1953), "On the Elements of Being I", *Review of Metaphysics* 7, pp. 3–18.

The following back volumes of ***Acta Philosophica Fennica*** are available from Bookstore Tiedekirja, Snellmaninkatu 13, FI-00170 Helsinki, Finland, tel. +358−9−635 177, email: tiedekirja@tsv.fi, www.tiedekirja.fi:

Fasc. LXXVII (2005): Ilkka Niiniluoto and Risto Vilkko (eds.): Philosophical Essays in Memoriam Georg Henrik von Wright. 167 pp.

Fasc. LXXVIII (2006): Tuomo Aho and Ahti-Veikko Pietarinen (eds.): Truth and Games – Essays in Honour of Gabriel Sandu. 322 pp.

Fasc. LXXIX (2006): Floora Ruokonen and Laura Werner (eds.): Visions of Value and Truth – Understanding Philosophy and Literature. 205 pp.

Fasc. LXXX (2006): Sami Pihlström (ed.): Wittgenstein and the Method of Philosophy. 239 pp.

Fasc. LXXXI (2007): Matti Häyry: Cloning, Selection, and Values. Essays on Bioethical Intuitions. 197 pp.

Fasc. LXXXII (2007): Juha Manninen and Ilkka Niiniluoto (eds.): The Philosophical Twentieth Century in Finland. A Bibliographical Guide. 468 pp.

Fasc. LXXXIII (2007): Juhana Lemetti and Eva Piirimäe (eds.): Human Nature as the Basis of Morality and Society in Early Modern Philosophy. 206 pp.

Fasc. LXXXIV (2008): Tim De Mey and Markku Keinänen (eds.): Problems from Armstrong. 219 pp.

Fasc. LXXXV (2008): Hanne Appelqvist: Wittgenstein and the Conditions of Musical Communication. 152 pp.

Fasc. LXXXVI (2009): Sami Pihlström and Henrik Rydenfelt (eds.): Pragmatist Perspectives. 295 pp.

Fasc. LXXXVII (2010): Virpi Mäkinen (ed.): The Nature of Rights: Moral and Political Aspects of Rights in Late Medieval and Early Modern Philosophy. 257 pp.

Fasc. LXXXVIII (2010): Leila Haaparanta (ed.): Rearticulations of Reason. Recent Currents. 274 pp.

Fasc. LXXXIX (2012): Ilkka Niiniluoto and Sami Pihlström (eds.): Reappraisals of Eino Kaila's Philosophy. 232 pp.

Fasc. XC (2013): Jaakko Hintikka (ed.): Open Problems of Epistemology – Problèmes ouverts en épistémologie. 207 pp.

Fasc. XCI (2015): Gabriel Sandu: Logic, Language and Games. 139 pp.

Fasc. XCII (2016): Georg Meggle and Risto Vilkko (eds.): Georg Henrik von Wright's Book of Friends. 250 pp.

Fasc. XCIII (2017): Ilkka Niiniluoto and Thomas Wallgren (eds.): On the Human Condition – Philosophical Essays in Honour of the Centennial Anniversary of Georg Henrik von Wright. 463 pp.

Action, Value and Metaphysics
Proceedings of the Philosophical Society of Finland Colloquium 2018

Edited by Jaakko Kuorikoski & Teemu Toppinen

The articles in this collection are based on presentations given at the 2018 colloquium of the Philosophical Society of Finland, held in Helsinki on 11-12 January 2018. This colloquium represented, in certain ways, a break from a long tradition. While the annual colloquiums of the society have, in the past, been built around a “one word” theme (e.g., truth or virtue), and included talks given only in Finnish or in Swedish, this time the objective was an open, general, and rigorously refereed conference in which philosophers could disseminate and discuss their best work. It was also decided that the conference presentations could also be given in English, not only to attract contributors beyond our borders, but most of all to better reflect and serve the diversifying Finnish philosophy. This collection is a result of an open call for papers and presents a sample of current philosophical work in Finland, with topics ranging from collective responsibility to philosophy of action and from metaphysics to metaethics.

JAAKKO KUORIKOSKI is an associate professor in New Social Research at the University of Tampere.

TEEMU TOPPINEN is a university researcher in practical philosophy at the University of Helsinki.

ISBN 978−951−9264−89−9
ISSN 0355−1792

Juvenes Print
Helsinki 2018