
NTU
NOTTINGHAM
LAW SCHOOL

NOTTINGHAM LAW JOURNAL

VOL. 28(2) **2020**

The *Nottingham Law Journal* is an independently refereed and indexed (*Hein Online*) journal, normally published in Summer each year. Contributions of articles, case notes and book reviews to the *Journal* are welcomed. Intending contributors are invited to contact the Editor for a copy of the style sheet, which gives details of the format which submissions must follow. Submissions and enquiries should be addressed to the Editor for academic year 2020–2021:

The Rev'd Dr Helen Hall, Nottingham Trent University, 50 Shakespeare Street, Nottingham, NG1 4FQ. Telephone 0115 848 2130. Dr Hall may be contacted on the following e-mail address: **Helen.hall@ntu.ac.uk**. Style notes and further details about the Journal are available on request.

Intending subscribers should please contact Miss Kerri Gilbert at the above address. Intending subscribers in North America are advised to contact Wm W Gaunt & Sons. Inc, Gaunt Building, 3011 Gulf Drive, Holmes Beach, Florida 3417 2199.

The citation for this issue is (2020) 28(2) Nott L J.
ISSN No. 0965–0660

NOTTINGHAM LAW JOURNAL

VOL. 28(2) **2020**

CONTENTS

NOTTINGHAM LAW JOURNAL

VOL. 28(2) 2020

EDITORIAL

This commitment of authors and colleagues to the Nottingham Law Journal is always impressive, but the response in the ill-stared year of 2020 has been truly humbling. It is a privilege to have edited such a high quality and wide-ranging collection of contributions amidst so many challenges. The journal has of course continued its commitment to peer review, and although for obvious reasons I am not at liberty to name the scholars who have assisted as peer reviewers, it would be remiss of me not to pay tribute to their dedication. At a moment in time when the academic community has been reacting to sudden and dramatic demands in order to simply keep universities functioning, a willingness to take on an extra burden is extremely laudable.

The process of peer review and revisions has been slower than usual, and I am also indebted to each one of the authors for their hard work, understanding and patience. I feel that we can be justly proud of the result. The edition reflects the consistent tradition of over forty years, of producing a general interest journal engaging with a broad range of topics. We have a reflection by Anahit Manasyan on "Democratic Ethics in the Modern World: Myth or Reality" a topic which goes to the foundation of contemporary legal and Constitutional frameworks. John Cheung reflects on the role of tort in protecting essential rights, providing a bridge between the worlds of public and private law in "Balancing Fundamental Rights in Private Law through the Double Proportionality Test".

We then step squarely into the realm of private law for several fascinating discussions. Jesus Ezurmendia examines "*Res Judicata*: A Gap Between Civil Litigation and Arbitration" a theme which is only likely to grow in importance in an increasingly global and complex world; whilst Juan Pablo Murga Fernandez considers "The Doctrine of Frustration in Spanish Law: Its Configuration in Light of the Pandemic", a study which only addresses matters of urgent present concern, but deeper questions which will apply even when the current crisis has abated.

I am of course indebted to all of these contributors, and enormously grateful to the editorial team, Daniel Gough as Deputy Editor, Linda Mururu as Postgraduate Associate Editor, and Selbi Durdiyeva as Acting Associate Editor. In addition the help of our administrative assistant Kerri Gilbert has, as always, been invaluable. I am also grateful for the advice and support offered by previous editors who remain as colleagues, Janice Denoncourt, Helen O'Nions and Tom Lewis.

THE REV'D DR HELEN HALL

NOTTINGHAM LAW JOURNAL

VOL. 28(2) 2020

ARTICLES

The address for submission of articles is given at the beginning of this issue.

DEMOCRATIC ETHICS IN THE MODERN WORLD: MYTH OR REALITY?

*ANAHIT MANASYAN**

ABSTRACT

The idea of democratic ethics is considered and presented in the article. According to the author, democratic ethics is a guiding philosophy, a set of principles and values, governing the exercise of the power of people and the state power, which sets the limits, beyond which the exercise of any power becomes inadmissible. The author presents the principle of "expedient self-restraint", which should underlie both behaviour of any individual, as well as political behaviour of the power and political actors. Analysing the idea of democracy, the author draws a conclusion that for the proper perception of the concept of democracy just the feature of the implementation of power by the people and for the people is not enough, and the doctrines of democracy need to be re-evaluated. Though democracy is a form of governance within the frames of which people are the source of power, the basis for this should be the human being as the highest value and the aim of guaranteeing possibilities for individual self-expression and self-realisation. Moreover, the power of people should not be an unlimited one. The axis of the representation element in the frames of the modern doctrine of democracy, in turn, should be the exercise of state power professionally. For the realisation of the discussed aim, election of representatives of power based on the idea of elitism is necessary, taking as a basis the features of competence and professionalism. Moreover, the principle of separation and balance of powers is necessary not for the final isolation of the branches of power from each other, but for ensuring their interaction and cooperation. The Armenian constitutional framework and its peculiarities are also studied in the context of the discussed issues. Special attention is paid to the study of democratic ethics in crisis and emergency situations, highlighting the importance of is observance.

Keywords: democracy; democratic ethics; expedient self-restraint; the power of people; limits of power; elitism; professionalism; human rights; individual self-expression and self-realisation; rethinking the principle of separation and balance of powers.

*Vice-Rector of the Academy of Justice of the Republic of Armenia, Associate Professor at the Chair of Constitutional Law of Yerevan State University, PHD in Law, Member of the Commission on Evaluation of the Activities of Justices (Republic of Armenia).

INTRODUCTION

"Democracy" is one of the widespread terms used in everyday life, and it is difficult to find anyone unaware of this term. At the same time, the mentioned notion is not always properly perceived and in some cases is realised in a distorted way, particularly when the authoritarian and totalitarian regimes declare themselves as democratic, invoking the support of people.[1] Whereas one should take into consideration that existence of expressing people's support phenomena, for instance, announced results of elections, etc., cannot be the sole evidence of the real power of people and democracy. Moreover, the power of people is often perceived as an unlimited one, and no attention is paid to the ethical rules, which should underlie all the democratic processes. Hence, democracy presupposes the existence of a number of circumstances, including democratic ethics, the essence of which we consider necessary to reveal in this context.

DEMOCRATIC ETHICS AS A BASIS FOR DEMOCRATIC PROCESSES

It is considered that the concept of democracy is mainly discussed for the emphasis of the fact that the power belongs to the people. At the same time, very little attention is paid to the study of the limits of this power and the ethical rules, which should underlie its exercise. Moreover, the ethics, underlying the exercise of the state power, is not paid proper and thorough attention either. While it is obvious that one cannot speak about proper democratic processes out of these ethical rules and the limits of power.

The concept of "democratic ethics" can sometimes be found in modern literature.[2] At the same time, it is almost not discussed from the viewpoint of the limits of the people's and the state powers. Moreover, there are no complex research and uniform academic approach on this concept.

In its dictionary definition, "ethics" is considered to be a set of moral principles**;** a theory or system of moral values; the principles of conduct governing an individual or a group; a guiding philosophy.[3] Hence, these moral principles and values are the limits, beyond which the conduct of an individual or a group becomes inadmissible.

Taking the above into account, we consider *"democratic ethics" as a guiding philosophy, a set of principles and values, governing the exercise of power of people and the state power, which sets the limits, beyond which the exercise of any power becomes inadmissible.*

Therefore, democratic ethics and the limits of the power should be the axis for the doctrine of democracy, its perception and revaluation. These issues will further be discussed in this article.

DOCTRINE OF DEMOCRACY IN MODERN CONSTITUTIONAL LAW AND THE NEED TO REVALUE IT

In its dictionary definition, "democracy" is considered to be the governance by people, where the supreme power belongs to the people and is implemented by the latter directly or by the representatives elected by it in conditions of the free electoral system. According to Abraham Lincoln, democracy is the power of people implemented by the

[1] Howard Cincotta (ed.), 'What Is Democracy?', <https://web-archive-2017.ait.org.tw/infousa/zhtw/DOCS/whatsdem/whatdm2.htm> accessed 4 May 2018.

[2] For instance, Paul Blokker, 'Democratic Ethics, Constitutional Dimensions, and Constitutionalisms' in A Febbrajo and W Sadurski (eds), *Central Eastern Europe after TransitionL Towards a New Socio-Legal Semantics* (Ashgate 2010).

[3] Dictionary by Merriam-Webster, <https://www.merriam-webster.com/dictionary/ethic> accessed 5 May 2020.

people and for the people.[4] In another definition, democracy is considered to be a form of governance, in conditions of which citizens implement the right to make (political) decisions personally or via representatives elected by them. Therefore, the circumstance of recognising people as the source of power underlies democracy.[5] Democracy is also presented as a state-political formation of society, where the people are the source of power.[6]

In literature, various circumstances are presented, presupposing the existence of the so-called "good, qualified" democracy, for instance, the existence of the universal suffrage, periodic, free, competitive, fair elections, more than one political party, more than one source of information, etc.[7]

Summarising the above views, we draw a conclusion that though each of these approaches touches upon the discussed issue from a different viewpoint, an important circumstance is crucial for all of them – the fact that the source of power in democratic social societies is the people. Hence, to our mind, the first key point for characterisation of the phenomenon "democracy" is the fact that in comparison with other forms of governance, in case of democracy, the real source of power is the people, the power belongs to the people, is implemented by the people and for the people.

A question arises in this context whether the above-mentioned universally recognised definition is sufficient for proper perception of the essence of democracy. Moreover, does this definition allow to speak about existence or absence of democracy in general or does it show the latter as just an aim, which should be sought, but will never be perfectly realised?

It is worth mentioning in this context that democracy is just a social idea, aim, and it is rather difficult or even impossible to thoroughly realise the latter. The proponents of the mentioned approach consider that in this case, the main methodological aspect should be the following: democracy should not be considered as a quality, which is typical or not for a particular society, but the main emphasis should be given to the circumstance that "the people" should themselves reply to the question how democratic it is, how democratic it should be and how democracy can be widened.[8]

In order to answer the mentioned questions, an analysis of several important issues is necessary, which will be touched upon below.

The view is widespread in the literature that democratic governance brings a number of problems, which make the effectiveness of the latter doubtful.

One of the most important negative phenomena, existing in democratic governance systems, is the formation of "mediocre leaders" or so-called "slaves of slogans", which, in turn, leads to the formation of low culture. According to Tocqueville, even if good leaders can come to power, they are not able to implement the long-term and society-wide projects due to being at the mercy of a fickle public with diverse interests, which is changing mandates for government with each election. This circumstance, in turn, leads to ineffectiveness of power.[9] Moreover, it is emphasised that in the mentioned

[4] Howard Cincotta (ed.), *'What* Is Democracy?', <https://web-archive-2017.ait.org.tw/infousa/zhtw/DOCS/whatsdem/whatdm2.htm> accessed 4 May 2018.

[5] Filosofiya: Entciklopedicheskii slovar, pod redakciey AA Ivina (Moskva, Gardariki, 2004) [A. A. Ivina (ed.), Philosophy: Encyclopedic Dictionary (Moscow, Gardariki, 2004)], <http://dic.academic.ru/dic.nsf/enc_philosophy/313> accessed 4 May 2018.

[6] Filosofskii enciklopedicheskii slovar [Encyclopedic Dictionary of Philosophy], <https://dic.academic.ru/dic.nsf/enc_philosophy/313/%D0%94%D0%95%D0%9C%D0%9E%D0%9A%D0%A0%D0%90%D0%A2%D0%98%D0%AF> accessed 4 May 2018.

[7] Leonardo Morlino, *Qualities of Democracy: How to Analyze Them* (Florence, Istituto Italiano di Scienze Umane 2009) 3.

[8] Frank Cunningham, *Theories of Democracy: A Critical Introduction* (London and New York, Routledge 2002) 144.

[9] *Ibid* 17.

systems the person, who is proficient in negotiations and "commercial" relations, hence, also in establishing ad hoc coalitions for support of concrete projects, becomes a successful politician.[10] Even guides on concepts and principles of democratic governance emphasise the self-advertisement function of representatives and stress that their success is possible just in conditions of realisation of this function, as a person cannot be famous if he/she does not advertise himself/herself.[11]

One of the most important shortcomings of democratic social systems is the syndrome of the so-called "empty space". The essence of the latter is in the fact that the carrier and source of power, that is – people, is a non-identifiable, abstract phenomenon, consisting of continuously changing masses. As a result, a so-called "empty space" is formed, in the context of which the power is implemented by elected individuals and people themselves are in fact alienated from power. Consequently, such power is immediately occupied by individuals, which play with public opinion or manipulate it. This circumstance, in turn, leads to demagogic governance of populist politicians or authoritarianism masked as democracy[12].

The next shortcoming of democracy is the fact that in the case of the latter, the state is continuously balancing between ochlocracy and dictatorship. Moreover, in this case, the tyranny of a person is substituted with the tyranny of the majority, which is not better, if not worse from the viewpoint of the normal course of social relations.[13] The fact that even the modern democracies cannot restrain from plutocracy is added to this.[14]

Irrationality is also one of the greatest shortcomings of democratic governance systems, the basis of which is the impossibility of the uninformed crowd to properly perceive its interests.[15]

In this context, the individual is the main subject of the research. The reason for that when the issue of "rationality" of a group is discussed, it is possible to consider it just from the aspect of the behaviour of the members of a group. The approach exists in the literature that the rational individual is also selfish. Moreover, it is emphasised that the political behaviour of people is similar to the behaviour of participants of economic relations; hence, it is presupposed that they take into account only their own interests. From this viewpoint, it is obvious that for instance, individuals pay for concrete goods more than necessary, noting not just the welfare of the manufacturer or salesperson. Taking the above into account, the presented authors consider an extension of the economic model of political behaviour crucial. In this context, it becomes obvious that when a person elects, runs for office or pursues certain policies when in office, this is assumed to be done not because of the concern for the public good. Moreover, it will also be impossible to guarantee that an individual will follow moral rules agreed on by the philosophers as being necessary for harmonious social life. This circumstance, in turn, makes crucial the issue that corresponding norms and institutions should be formed, taking into account the continuous possibility of the existence of selfish behaviour, and channelling it in such a direction that it becomes beneficial rather than detrimental

[10] Jack L Walker, 'A Critique of the Elitist Theory of Democracy' (1966) 60 The American Political Science Review 285, 291.

[11] Mathias Kamp (ed.), *Concepts and Principles of Democratic Governance and Accountability*, (Konrad-Adenauer-Stiftung, 2011) 18–19.

[12] Cunningham (n viii) 179.

[13] The most important issue here is the fact that nobody can guarantee that the choice of the society is correct. For instance, Hitler gained power in a democratic way, getting majority in Reichstag (Kamil Galeev, 'Monarhiya i Respublika: dostoinstvo i nedostatki' [Kamil Galeyev, 'Monarchy and Republic: Advantages and Disadvantages'], <http://schools.keldysh.ru/labmro/lib/projects/ galeevk.htm> accessed 4 May 2018).

[14] *Ibid*.

[15] Cunningham (n viii) 22.

to the interests of the members of the society and normal for development of social relations. In economic relations, the mentioned result is in some sense also guaranteed due to the fact that they are premised on the assumption of purely self-regarding behaviour.[16]

In this context, the perception and analysis of the behaviour of political parties are also highlighted. Parties are composed of political actors, who seek power not in order to implement favoured policies, but in order to attain income, prestige and power. For this purpose, they join others in political parties to compete for power. As a result, it becomes evident that in democratic political systems, parties are equivalent to entrepreneurs in economic relations. Hence, political parties form and implement such a policy, which, in their opinion, can guarantee more votes just as the entrepreneurs produce products they believe will gain the most profits. It is evident that the state power should implement certain social functions, and the governing party should keep enough voters sufficiently satisfied to be re-elected. In other words, in the mentioned "exchange relations" citizens should get something in return for their votes. But these benefits are by-products of the motivating goal of getting elected and staying in power in the same way as providing a customer with a functioning car is a by-product of a dealer's effort to make a sale.

Realising this, parties publicise their ideologies as advertisements. As the aim of a party is just to be elected, its leaders do not care about the intrinsic value of ideology but develop one they think will attract the largest number of voters.[17]

Taking the above into consideration, we draw a conclusion that the shortcomings of democracy and the distortions, emerging in social systems in the result of them, are almost identical with the shortcomings of other types of state regimes and distortions, emerging in the context of the latter. For instance, shortcomings of absolute monarchy, which is considered to be an expression of non-democratic political regime, are the following: dependence of the destiny of the whole state on the will and qualities of one person; conflicts with regard to the inheritance of the throne; the fact that the main means for overcoming dissatisfaction of people becomes overthrowing the power, which can lead to civil wars, etc.[18] It is obvious that from the viewpoint of results, the discussed shortcomings are almost identical with the shortcomings of democracy and distortions of democratic social systems. Taking the above-mentioned arguments into account, some of the literature demonstrates that in fact in the situation when the term "democracy" has a potential to meaning everything to everyone, there is a danger that term democracy then means nothing at all.[19]

We believe that the idea of democracy can be transformed to a senseless concept, and the same problematic situations typical for other types of state regime can emerge in one situation – in case of considering democracy just as a state regime within the frames of which power belongs to the people.

Hence, in our opinion, the existing perceptions and doctrinal approaches concerning the essence of democracy should be reevaluated. Moreover, invoking the above-presented and universally recognised definition, is not sufficient for the proper perception of the essence of democracy; and in this context, several other circumstances should be taken into account.

[16] *Ibid* 101–110.

[17] *Ibid* 101–110.

[18] Kamil Galeev, 'Monarhiya i Respublika: dostoinstvo i nedostatki'[Kamil Galeyev, 'Monarchy and Republic: Advantages and Disadvantages'], <http://schools.keldysh.ru/labmro/lib/projects/ galeevk.htm> accessed 4 May 2018.

[19] Mathias Kamp, *Concepts and Principles of Democratic Governance and Accountability* (Konrad-Adenauer-Stiftung, 2011) 10.

From this viewpoint, it is crucial that the approach of recognising people as a source of power is not an end in itself and can never be perceived as such. Moreover, we think that the reasons for acknowledging this approach are natural and are based on noting the nature of the human being and peculiarities of the formation of societies. From this aspect, it is obvious that any social-political formation is established in order to firstly satisfy the individual's needs and preserve the existence of the human being. Hence, it is also evident that the best means for reaching the mentioned aim should be the recognition of people as a source of power.

This point, in turn, presupposes *that the doctrine of democracy should itself be based on the logic that the human being is its axis, for whom possibilities of self-expression and self-realisation shall be guaranteed, it should also be based taking into consideration the individual value of each human being.* We consider that the above-mentioned circumstance is the essence, main ideal, the goal of democracy, its key characteristic and peculiarity, distinguishing the latter from other forms of state-political formation of the society. Hence, democracy is not an end in itself, and its value is not in the idea that power belongs to the people, but in the factual realisation of the above-mentioned aim.

Summarising the presented analysis, we draw a conclusion that the existence of democracy is conditioned by the extent the discussed aim has been realised in reality. This, in turn, presupposes that all the other features presented with regard to the mentioned type of state regime emanate from the above-noted peculiarity, are aimed at the realisation of this goal, which assumes that conditioned with various developments in social life and its perception, these features can be changed over time.[20]

We would also like to touch upon the Armenian context with this regard. The Constitution of the Republic of Armenia, adopted in 1995, played a significant role in the establishment of democracy in the Republic of Armenia, strengthening the bases of a rule-of-law State, finding constitutional solutions in crisis situations, the gradual development of the institutions of the state power, and prescribing the constitutional safeguards for the protection of human rights. Moreover, in the Republic of Armenia, as in other newly independent states, the constitutional solutions have been built upon the objective of forming the public authorities of an independent state and safeguarding their performance. At the same time, the constitutional-legal model adopted in the Constitution of the Republic of Armenia in 1995 was predominantly power-centred. There was no clear attitude towards the constitutional recognition and stipulation of the human being as the highest value. Moreover, human dignity was not enshrined and perceived as an object for protection under constitutional law. The approach typical to the prior Soviet legal system with respect to this issue was still being applied, and human dignity was considered just an object of protection within the scope of criminal law legislation.

The further process of enrooting of democracy and universal values showed the necessity of carrying out constitutional developments with this regard. Subsequently, Article 3 of the Constitution of the Republic of Armenia in the edition of 2005 enshrined that the human being, his/her dignity, fundamental rights and freedoms represent the highest values the state must protect. The State shall ensure the protection of fundamental human and citizen's rights and freedoms, in conformity with the principles and norms of international law. The State shall be bound by fundamental human and citizen's rights and freedoms as directly applicable law.

This was a visible development towards the adoption of a human-centred constitutional model, as well as constitutional recognition and stipulation of mechanisms

[20] Anahit Manasyan, *Constitutional Stability as an Important Prerequisite for Stable Democracy*, (Yerevan, 2020) 9–21.

concerning this. Anyway, notwithstanding some achieved progress with this regard, the necessary thorough constitutional prerequisites for a more consistent implementation of the rule of law principle and enrooting of democracy were still in need of development. This was the reason that the strengthening of constitutional safeguards for guaranteeing, ensuring and protecting the basic human rights, the clarification of the framework of possible limitations of these rights, were regarded as an important direction of 2015 constitutional reforms of the Republic of Armenia. It was stipulated in the Concept Paper on the Constitutional Reforms of the Republic of Armenia that "This requires creating sufficient and necessary constitutional-legal prerequisites for implementing the principle of the rule of law, which necessarily requires: safeguarding the direct application of basic rights in systemic integrity, strengthening the constitutional foundation for their protection, and clarifying the positive obligation and public-legal responsibility of the state in this matter; and confining the discretion of the power by law. The main principle of this approach is that the safeguarded protection of rights ensures the democratic freedoms and the direct application of rights, and limits discretion and manifestations of the subjectivity of the power".[21]

Moreover, the Constitution of the Republic of Armenia in the edition of 2015, in comparison with constitutional regulations of 2005, according to which the *state* was bound by fundamental human and citizen's rights and freedoms as directly applicable law,[22] defines the requirement of being bound by fundamental human and citizen's rights for the *public power.*[23] We consider that the mentioned amendment is of conceptual nature and crucially transforms the essence of the discussed phenomenon. This is firstly emphasised from the aspect that in this context, the requirement of being bound by fundamental human and citizen's rights is also extended to the power of people, as the idea of public power includes not just state, but also people's power.[24] Moreover, the Concept Paper on the Constitutional Reforms of the Republic of Armenia directly emphasises that the conceptual approach is that when administering power, *the people and the State* must be limited to the basic human and citizen's rights and freedoms.[25] Hence, the power of people and its realisation cannot be considered as an end in itself either and are in any case bound by fundamental human and citizen's rights and freedoms.[26] Moreover, according to the Armenian Constitution in the edition of 2015, the human being shall be the highest value in the Republic of Armenia. The inalienable dignity of the human being shall constitute the integral basis of his or her rights and freedoms.

At the same time, though constitutional recognition and stipulation of the mentioned ideas is of conceptual importance, enrooting democracy and the rule of law is firstly conditioned by the extent the discussed ideas and aims have been realised in reality.

[21] Specialized Commission on Constitutional Reforms adjunct to the President of the Republic of Armenia, *Concept Paper on the Constitutional Reforms of the Republic of Armenia* (Yerevan, September 2014), <http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2014)033-e> accessed 16 June 2018.

[22] Article 3 of the Constitution of the Republic of Armenia in edition of 27 November, 2005.

[23] Article 3 of the Constitution of the Republic of Armenia in edition of 6 December 2015.

[24] Public power consists of the following three structural elements: power of people, state power, local self-governance power (Suren Avakyan, 'Publichnaya vlast: konstitucionno-pravovye aspekty; (2009) 2 Vestnik Tyumenskogo gosudarstvennogo universiteta [Suren Avakyan, 'Public Power: Constitutional-Legal Aspects' (2009) 2 Bulletin of the Tyumen State University] 5, 6).

[25] Specialized Commission on Constitutional Reforms adjunct to the President of the Republic of Armenia, *Concept Paper on the Constitutional Reforms of the Republic of Armenia* (Yerevan, September 2014), <http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2014)033-e> accessed 16 June 2018.

[26] In this context the example of the Republic of Georgia is interesting. Article 7 of the Constitution of the latter defines that while exercising authority, the people and the State shall be bound by the universally recognized human rights and freedoms as directly applicable law (<http://www.constcourt.ge/en/ court/legislation/constitution.page> accessed 4 May 2018).

Hence, our further steps should be directed towards the factual implementation of those ideas in practice.

Summarising the above, we consider that *the key feature for democracy is the circumstance that in conditions of the latter people are the source of power, power is exercised by the people and for the people, having as a basis the human being as the highest value and having an aim to guarantee for the possibilities of self-realisation and self-expression for an individual.*[27]

Hence, we draw several other important conclusions presented below:

It is obvious that distortions of social systems, resulting from shortcomings typical for the rule of people, have entailed to such a situation that the aim, for which democracy as an idea, ideal was established, is also distorted. Moreover, institutions typical for the rule of people often start to serve opposite to the mentioned aim.

Hence, the main conclusion here should be the following: institutions formed in order to regulate social relations should be implemented just for the goals they were formed for. Otherwise, notwithstanding the fact of observance of various formal requirements, the essence of the institution is distorted, which, in turn, makes the necessity of its existence senseless.

The next important conclusion in the mentioned context is the following: in order to realise the above-mentioned aims and avoid the listed shortcomings, adequate constitutional and political culture is necessary. Hence, the existence of features typical for democracy is possible just in conditions of adequate pre-democratic constitutional and political culture. Democracy is not a peculiarity, originally belonging to state and society, and can be reached just at a certain stage of social development. Moreover, it is obvious that even in case of realisation of the mentioned aim and existence of democracy, the latter will periodically come to the above-listed distortions, and will result in the factual absence of democracy as a constitutional-legal feature of the state. Hence, we consider that development and strengthening of a democratic state is a continuous circle – a transition from a pre-democratic social system to a democratic one and vice versa. Moreover, during the whole period, preceding and succeeding democratic social systems, the process of formation of an adequate constitutional and political culture goes on. This is the reason why, according to many authors, democracy shall be considered not just as a state, but also as a process, and from this viewpoint terms "democracy" and "democratisation" should be perceived as synonyms.[28]

Hence, there is an interesting fact in this context: the existence of other types of state regime and phenomena typical for them becomes an intermediate mechanism for the formation of democratic social system and culture. Moreover, one can conclude from the presented analysis that it is impossible to avoid this.

The above leads to a conclusion that the theory of democracy is nothing else than a means aimed at guaranteeing the key circumstance necessary for harmonious existence of social systems, that is – the human being is the highest value, possibilities for self-realisation and self-expression for an individual are guaranteed, and the state recognises the inherent value of a human being. Moreover, as it was already mentioned, the approach of recognising people as a source of power is not an end in itself, the

[27] In the mentioned context the view presented by certain authors is worth mentioning, according to which, from the "moral" viewpoint the essence of democracy is in the development of all the social capacities for each concrete member of society, and the "main criteria" of democracy is providing individuals with equal effective right to live as fully as they may wish (Cunningham (n viii) 149).

[28] See Nikolay Baranov, 'Istoricheskii obzor razvitiya demokratii'[Nikolay Baranov, 'Historical Review of the Development of Democracy'], <http://nicbar.ru/ theoria_democraty1.htm> accessed 4 May 2018).

reasons for adopting this approach are natural and are based on taking the nature of the human being and peculiarities of formation of societies into account. Though distortions of social systems, resulting from the rule of people, are almost identical with shortcomings of other types of state regime,[29] at the same time, it is obvious that any social-political structure is formed first to satisfy the requirements of an individual and to preserve the existence of the human being. Hence, it is evident that for individuals the best means for reaching the mentioned aim can be recognising themselves as the source of power. Consequently, we draw a conclusion that recognising people as a source of power is not an end in itself, is just a more acceptable means for reaching the key goal of societies from the viewpoint of the human being's individual and social peculiarities. Summarising the above, it should be noted that *one can speak about the existence of democracy just in the situation when each individual feels that in the concrete social system the human being is the highest value, possibilities of his/her self-realisation and self-expression are guaranteed, as well as an individual is valued.*

Moreover, handing the power to the people can serve the presented aim of social societies in case of existence of an adequate level of constitutional and political culture and in some sense is a result of its existence.

It is noted that formation of an adequate constitutional and political culture presupposes the need to establish and strengthen the ethical rules, principles and values, which should govern the exercise of any power. The latter should become the "red line", beyond which the exercise of power becomes inadmissible. It is considerable that in the contemporary world such values are expressed, as a rule, in Constitutions or relevant acts. Hence, their observations and protection become axial for the existence of democratic ethics and democracy. At the same time, even if Constitutions set rules, values and principles adequate for having democratic ethics and democracy, it is important to observe them in reality and have mechanisms for their protection in case they are violated. It is considerable that "ethics" in its dictionary definition is presented as a set of *moral* principals, a theory or system of *moral* values.

Hence, the most important phenomenon here is the principle of "expedient self-restraint", which, in our opinion, should underlie both behaviour of any individual, as well as the political beaviour of the power and political actors. This principle presupposes that people and political power should firstly restrain themselves by the discussed principles, values and rules. Without such self-restraint it will be impossible to speak about democratic ethics, hence also, real democracy, even in case written constitutional principles and rules, as well as their protection mechanisms exist.

As it was mentioned in the Concept Paper on the 2015 Constitutional Reforms of the Republic of Armenia, "The gap between the Constitution and real life should be overcome in principle. Constitutionality and the constitutioanlisation of social relations should help to find legal safeguards for solving political, social, economic and other issues, based on the fundamental truth that overcoming the deficit of constitutionality is the guarantee of stable development. In a state that ensures the rule of law, the political, economic and administrative potentials may not become integrated. The constitutionality of the goals and activities of political forces must be guaranteed and secured. The Constitution should underline the importance of constitutionalising the individual's

[29] With this regard Aristotle's viewpoint is worth mentioning, according to which majority can make good rulers not because any ordinary person is wise, but because of the pooled experiences and knowledge of many individuals. In other words the latter proposes the idea of quantitative wisdom. In this context "jury theorem" of Condorcet is also worth mentioning, according to which on the assumptions there is some decision that will be objectively the best for a society and each of the society's voters has a better than 50 percent chance of selecting it, the larger the majority votes for a particular option the more likely it is the best option voted for (Cunningham (n viii) 154).

social conduct and the government's political and public conduct, both of which should be based on the principle of the rule of law.[30]

The presented circumstances are the basic principles and values, which should guide the exercise of both the people's and the state power and form democratic ethics necessary for the proper perception of the doctrine of democracy.

THE IDEA OF "ELITISM" AND ITS ROLE IN EFFECTIVE EXERCISE OF THE POWER OF PEOPLE

In this context, it is important to note that besides the traditional "majoritarian" theory of democracy, there are also other theories, which conceptually differ from the first one on a number of essential issues.

One should take into account that they also (for instance, pluralist, deliberative and other theories) suggest various ways for overcoming distortions, evolving in social systems. Anyway, to our mind, none of them can surmount the main reason for the mentioned shortcomings, that is – individual and social peculiarities of the human being, particularly, his/her selfish nature, as a result of which the mentioned main distortions take place.

In this context, we consider necessary to reveal and analyse the key peculiarities of one of the democratic theories – the elite theory of democracy.

The founder of the elite theory of democracy is Joseph Schumpeter, who insisted that democracy does not mean that people directly govern. Democracy just means that people have an opportunity to accept or reject individuals, who will govern them. The popularity of this method is conditioned by the fact that the candidate leaders have the possibility of free competition for getting the votes of the electors.

The proponents of the elite model of democracy separate the society to the governing minority (elite) and not governing majority (masses). Masses are not interested in politics, do not have enough knowledge and thorough information, cannot make proper decisions, because of which they voluntarily transfer the right to manage political processes to the elite. As the majority of citizens are irrational, are not competent and have non-stable preferences, political participation of masses are limited to elections.[31] Moreover, according to D.S. Mill, the main option to solve the mentioned issue typical for the majority of citizens is transferring the power to educated individuals to guarantee the thoughtful choice of leaders and policies, at the same time, encouraging people's general participatory democracy via transferring practical knowledge on reasonable self-governance to them. Mill also considers it necessary to develop a "political morality", which will give representatives wide discretion to act in a way, which they think will be in compliance with the best interests of their voters. At the same time, determining proper scope for political representatives becomes crucial in this context. As a criterion for the concretisation of the mentioned limits, it is proposed that representatives are free in implementation of their discretion, but they should exercise it in the interests of their voters, who are also thought to be capable of independent judgement and are not just charged on representatives to be taken care of.[32]

[30] Specialized Commission on Constitutional Reforms adjunct to the President of the Republic of Armenia, *Concept Paper on the Constitutional Reforms of the Republic of Armenia* (Yerevan, September 2014). <http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2014)033-e> accessed 16 June 2018.

[31] Nikolay Baranov, 'Sovremennaya demokratiya: evolyucionnyi podpod'[Nikolay Baranov, 'Modern Democracy: Evolutionary Approach'], <http://textbooks.studio/uchebnik-teoriya-politiki/elitarnaya-model-demokratii-23148.html> accessed 4 May 2018.

[32] Cunningham (n viii) 92–93.

Taking into consideration the necessity of professional governance and highlighting its significance from the aspect of the realisation of the main aim of social systems, at the same time, we consider the critical approach of proponents of elite democracy, according to which power of people should be restricted by the possibility of accepting or not accepting its rulers, to be extreme. Such a solution does not overcome the discussed problems, emanating from the individual nature of the human being. Hence, as a result, the presented threat of distortion of social systems does not disappear.

Therefore, the main conclusion here is the following: marginal approaches, underling both majority and elite theories of democracy, do not give an opportunity to overcome distortions of social systems. At the same time, each of them propose certain significant and key viewpoints, which have exceptional importance from the aspect of perfection of the theory of democracy. Hence, we consider that solutions necessary from the viewpoint of regulation of social relations and development of social systems can be found just in case of combining the presented approaches and this should underly the essence of the doctrine of democracy.

Particularly, as already noted, implementation of the majority power should be based on the human being as the highest value, the aim of this power should be guaranteeing possibilities for self-expression and self-realisation for an individual, the power of the majority, in turn, should be restricted by the mentioned circumstances and by fundamental human rights and freedoms.

At the same time, the idea of *professional* governance proposed by the elite democracy theory has exceptional importance from the viewpoint of finding adequate solutions to various arising problems, as well as from the aspect of increasing effectiveness of the governance system, and to our mind, it should also underly the theory of democracy and lay at the essence of democratic ethics.

In this regard, it is worth mentioning that the representation element is one of the mandatory peculiarities of the modern doctrine of democracy and was considered by philosophers as an opportunity provided to democracy to gain new form and quality. At the same time, the axis of the latter should be not just the exercise of power by state and local self-governance bodies and their officials, but the exercise of power by them *professionally*. In this regard, it is important that gaining power should not be an end in itself for the people, exercising power. They should be endowed with high professionalism, notwithstanding the fact whether they implement power authorities or not. Moreover, on the basis of implementation of these authorities should be the human being as the highest value and the aim of the power should be guaranteeing possibilities for self-expression and self-realisation for an individual, as well as the inherent value of every human being. The power, in turn, should be restricted by the mentioned circumstances and fundamental human rights and freedoms. Just in this situation, it is possible to overcome the fear that inclusion of representation elements lead to a number of internal difficulties, in particular, these institutions alienate the power from the demos in such a way that the question of critics to what extent it is legitimate to name the new system "democracy" is absolutely appropriate.[33]

With regard to the above-mentioned issue, it is necessary to analyse key ideas existent in literature on the formation and transformation of governing elite. In the theory of political science, two main approaches are presented on the mentioned issue – elitism and elitarism. The essence of elitism is the following: the election of the people, exercising power, is implemented on the basis of meritocratic principle (literally – "power

[33] Nikolay Baranov, 'Istoricheskii obzor razvitiya demokratii' [Nikolay Baranov, 'Historical Review of the Development of Democracy'], <http://nicbar.ru/ theoria_democraty1.htm> accessed 4 May 2018.

based on merit", from Latin *mereō* – merit, from Greek κράτος – power, governance). The essence of the latter is in the election of the best professionals, such people, who are more competent in the field of governing political processes. As a result of such an approach, the elite is formed on the basis of open competition – via the election of the best professionals and rulers. Exercise of power by them is perceived as a function of distribution of the managerial work and not as the main aim of career ambitions. In the mentioned situation, power functions belong to more prepared and talented representatives of society, for whom obtainment of power is not an end in itself. Moreover, they are endowed with high professionalism, notwithstanding the fact of whether they implement power authorities or not.

In comparison with the above, elitarism is based on the idea of separating the elite from masses, as a result of which the elite is considered to be a closed class, layer. In this case, formation of the elite is implemented in a way, excluding competition – via inheritance, class privileges, party or nomenclature belonging, etc. Power resources become accessible for the representatives of the elite, notwithstanding their abilities and governance skills. They become carriers of power functions due to belonging to the governing elite. Obtainment of power becomes the essence and the main aim of their political activities, and possession of ruling position in the governing hierarchy is dictated just by subjective motives (firstly by vanity and the need to satisfy ambitions) and not by social or state interests. Such a model of formation of the elite develops a ruling group, consisting of rulers, who have just career aims and are projected with the goal of preserving their corporate supremacy as long as possible.

In this context, it should also be noted that contradistinction of two qualities more required in the field of political governance, that is – loyalty and competence, is presented in the theory of management. As a result, each leader stands before a dilemma: in case of selecting candidates, who are just loyal to him/her, the system becomes ineffective; in case of selecting competent candidates, one deals with highly professional and talented individuals, as a result of which the authority of the ruler (head) as a manager can be threatened.[34]

Noting the above, we draw a conclusion that election of representatives of power on the basis of the idea of elitism, taking as a basis the features of competence and professionalism, is necessary for reaching the discussed aim of democracy and development of proper democratic ethics.[35]

In this context, we would also like to consider the doctrine of separation and balance of powers, which, in our opinion, also needs re-evaluation.

We particularly consider that this principle is necessary not for the final isolation of the branches of power from each other, but for *ensuring their interaction and cooperation*. Moreover, *the main aim of the principle of separation and balance of powers is not the existence of mechanisms of mutual control of state bodies, but firstly, finding more acceptable and effective joint solutions to problems*. Not by chance, the view is widespread in legal literature that just the separation of the state power allows their cooperation.[36] In this context, the legal position of the Federal Constitutional Court of Germany is

[34] Gennadiy Krasnolutskii, 'Politicheskaya alternative Rossii: elitism ili eletarism?' (2011) 1 Управленческое консультирование [Gennadiy Krasnolutskiy, 'Russia's Political Alternative: Elitism or Elitarism?' (2011) 1 Administrative Consulting] 89, 89–92.

[35] Anahit Manasyan *Constitutional Stability as an Important Prerequisite for Stable Democracy*, (Yerevan, 2020) 21–26.

[36] Dmitri Pulbere, 'Soblyudeniye prav cheloveka i control konstitucionnosti zakonov v obshem primere Respubliki Moldova' (2011) 1(51) Konstitucionnoye pravosudiye: Vestnik Konferencii organov konstitucionnogo kontrolya stran molodoi demokratii [Dmitriy Pulbere, 'Respect for Human Rights and Control of Constitutionality of Laws in General and on the Experience of the Republic of Moldova' (2011) 1(51) Constitutional Justice: Bulletin of the Conference of Constitutional Control Bodies of Young Democracies] 58, 70.

worth mentioning, according to which coordination of activities of constitutional bodies requires as a guarantee of their independence, that they harmoniously cooperate with each other while exercising their constitutional activities and deny such actions, which can cause harm to the authority of other constitutional bodies, thus endangering the Constitution itself.[37] The Constitutional Court of the Russian Federation also expressed a legal position, according to which separation of powers presupposes such a system of legal guarantees, checks and balances, which excludes the concentration of power in the hands of one of them, ensures the substantive and independent exercise of all the branches of power and simultaneously, cooperation.[38]

The above leads to a conclusion that in conditions of such a perception of the principle of separation and balance of powers, no branch of power can be considered as primary or secondary in comparison with other branches. Moreover, the main aim of the activities of all the branches of power should be finding more acceptable and effective joint solutions to the existing problems, and not implementing control over other power bodies as an end in itself.

It was stipulated in the Concept Paper on the 2015 Constitutional Reforms of the Republic of Armenia that "It is necessary, in the context of systemic integrity, to implement the constitutional principle of the separation and balance of powers more consistently, to guarantee harmony in the function-institution-power chain, to balance the functional, the mutually-balancing and mutually-checking powers of government bodies, and to strengthen the proper efficiency and functional independence of the various branches of power. The necessary constitutional prerequisites should be created for overcoming expressions of shady relations and subjectivity in the performance of state power functions, as well as safeguarding public-legal accountability and programmatic and goal-oriented activities of the state power". [39]

It is of conceptual importance in this context that the constitutional disputes between constitutional bodies must have clear resolution mechanisms. It is worth mentioning with this regard that as the result of 2015 constitutional reforms of the Republic of Armenia a new power of the Constitutional Court was enshrined, according to which the Constitutional Court shall settle disputes arising between constitutional bodies with respect to the constitutional powers thereof. This is a very important constitutional development, as it gives an opportunity to solve the constitutional disputes (via interpreting the Constitution) in a civilised, constitutional way and not just in the political dimension, "behind the curtains and scenes". According to Article 75 of the Constitutional Law of the Republic of Armenia "On the Constitutional Court", Constitutional Court, interpreting the corresponding provision of the Constitution, makes one of the following decisions on disputes between constitutional bodies with regard to their constitutional powers: 1) finds the action or inaction or the legal act of the respondent implemented/adopted by the exercise of the challenged power in conformity with the Constitution, 2) finds the action or inaction of the respondent implemented by the exercise of the

[37] Otto Luhterandt, 'Znacheniye reshenii Federalnogo Konstitucionnogo Suda dlya pravovogo poryadka Germanii' (2011, Erevan) Almanah (Konstitucionnoye pravosudie v novom tysyacheletii) [Otto Lukhterhandt, 'Role of the Federal Constitutional Court Decisions for the Legal Order of Germany' (2011, Yerevan) Almanac (Constitutional Justice in the New Millenium)] 123, 128.

[38] Valeriy Zorkin, 'Princip razdeleniya vlastey v deyatelnosti Konstitucionnogo suda Rossiiskoi Federatsii' (2008) 2(40)-3(41) Konstitucionnoye pravosudie: Vestnik Konferencii organov konstitucionnogo kontrolya stran molodio demokratii [Valeriy Zorkin, 'Principle of Separation of Powers in the Activities of the Constitutional Court of the Russian Federation' (2008) 2(40)-3(41) Constitutional Justice: Bulletin of the Conference of Constitutional Control Bodies of Young Democracies] 26, 28.

[39] Specialised Commission on Constitutional Reforms adjunct to the President of the Republic of Armenia, *Concept Paper on the Constitutional Reforms of the Republic of Armenia* (Yerevan, September 2014), <http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2014)033-e> accessed 16 June 2018.

challenged power in non-conformity with the Constitution, and the adopted legal act fully or partially invalid and in non-conformity with the Constitution, 3) finds the constitutional body – not a party of the proceedings, as being authorised to implement the challenged power, and dismisses the case on this basis.

In my opinion, the discussed institute can be an effective tool for revaluing the principle of separation and balance of powers in the above-presented context and developing the principle of mutual respect and cooperation between the constitutional bodies.

Summarising the above, it is noted that democracy is impossible without professional governance, proper separation and balance of powers, which, inter alia, comprise the democratic ethics.

DEMOCRATIC ETHICS AND THE WORLD NOWADAYS

We very often speak about the fact that democracy, hence, democratic ethics are just mythical ideas and not really existing phenomena.

At the same time, many studies point out the rise of democracy in recent years, highlighting that though there are flaws in democratic practices everywhere, the world is, on average, more democratic than ever before.[40]

We already noted the main factors, principles and values, which should underlie democratic ethics and in which conditions only it is possible to speak about the real democracy. We believe that their observance can become a reality and ensure the existence of democratic ethics. The rise and strengthening of democratic values in various regions and countries testifies this. At the same time, in all those countries this is a result of the proper level of constitutional and political culture, observance of relevant values and rules. Hence, the most important issue in this context is the proper perception of democratic ethics and democracy, as well as factual observance of the discussed ethical rules, principles and values. It is considerable that though we always speak about the importance of democracy and democratic ethics, they are very often endangered in crisis situations. From this viewpoint observance of democratic values and principles has become particularly important since the spread of COVID-19 pandemic last year. Many Council of Europe member states declared full or partial state of emergency. This situation lead to various unprecedented decisions, including massive restrictions of several fundamental rights, such as freedom of speech, freedom of assembly, freedom of movement, etc. Some states banned the publication of information related to COVID-19 from non-official sources, imposed severe restrictions on the publication of misinformation or the publication of information deemed to be "false", banned the possibility to assemble during the state of emergency or limited the freedom of assembly in other way, etc. Such restrictions were massive and had an enormous impact on global state of democracy. The reason is that democratic ethics and democracy are directly conditioned by the level of respect of human rights, hence also by legitimacy of restriction of those rights, which, in turn, depends on professional governance and proper separation and balance of powers. It is obvious that now we are facing an unprecedented situation, which very often requires prompt reactions. At the same time, this does not give an opportunity to restrict the fundamental rights, ignoring the criteria necessary for making the process legitimate.

Any crisis or an emergency situation is difficult, requires quick reactions. The rules, which should be followed in those situations, were developed with the aim of eliminating

[40] Department of Political Science, Aarhus University, '*The Rise of Modern Democracy*', <https://ps.au.dk/en/research/research-projects/the-rise-of-modern-democracy/> accessed 24 January 2021.

violations in such processes. Hence, they should be strictly followed even in such an unprecedented situation, which we have been facing since last year.

The history also testifies that during emergencies the abuse of human rights is the greatest. The Venice Commission emphasizes with this regard that emergency measures should respect certain general principles which aim to minimize the damage to fundamental rights, democracy and rule of law. The measures should be subject to the general conditions of necessity, proportionality and temporariness. The Commission highlights that even in a state of public emergency the fundamental principle of the rule of law must prevail. Hence, even in such situations limitations of fundamental rights should be subject to a triple test of legality (are prescribed by law), legitimacy (pursue a legitimate aim) and necessity (are needed to reach the aim and proportionate to it). It is considerable that the Venice Commission refers also to the proper implementation of the principle of separation and balance of powers, noting that "As a state of emergency involves derogations from the ordinary rules on distribution of powers, it is important, for the crisis management to be effective and coordinated and for the sake of equality and fairness of treatment of all citizens, that all state, regional and local institutions and bodies respect the principle of loyal cooperation and mutual respect between them".[41]

Therefore, democratic ethics is a particular necessity in our nowaday's world, needs a strict observance for protecting democracy in general and should be paid special attention in crisis and emergency situations.

CONCLUSION

Summarising the above, we consider that democratic ethics is a guiding philosophy, a set of principles and values, governing the exercise of the power of people and the state power, which sets the limits, beyond which the exercise of any power becomes inadmissible.

"Ethics" is firstly connected with *moral* principles and values. Hence, the most important phenomenon for democracy is the principle of "expedient self-restraint", which should underlie both behaviour of any individual, as well as political behaviour of the power and political actors. This principle presupposes that people and political power should firstly restrain themselves by the discussed principles, values and rules. Without such a self-restraint it will be impossible to speak about democratic ethics, hence also, real democracy, even in cases when written constitutional principles and rules, as well as their protection mechanisms, exist.

Hence, the key feature for democracy is the circumstance that in conditions of the latter people is the source of power, power is exercised by the people and for the people, having as a basis the human being as the highest value and having an aim to guarantee possibilities of individual self-realisation and self-expression, as well as to value the human being.

Moreover, though the representation element is one of the mandatory peculiarities of the modern doctrine of democracy and was considered by philosophers as an opportunity provided to democracy to gain new form and quality, the axis of the latter should not be just the exercise of power by state and local self-governance bodies and their officials, but the exercise of power *professionally*. From this viewpoint, it is important that obtainment of power should not be an end in itself for the people, exercising power.

[41] European Commission for Democracy Through Law (Venice Commission), *CDL-PI(2020)005rev-e Report – Respect for Democracy Human Rights and Rule of Law during States of Emergency – Reflections* (Strasbourg, May 2020), <https://www.venice.coe.int/webforms/documents/?pdf=CDL-PI(2020)005rev-e> accessed 24 January 2021.

Moreover, the basis of implementation of power authorities should be the human being as the highest value, and the aim of the power should be guaranteeing possibilities for individual self-expression and self-realisation, as well as the inherent value of the human being. The power, in turn, should be restricted by the mentioned circumstances and fundamental human rights and freedoms.

Hence, the features of democracy can undergo changes over time. Simultaneously, in order to characterise the social system as democratic, the key peculiarity of the latter should necessarily stay unchangeable, that is – the exercise of power will not be an end in itself, it shall be based on the human being as the highest value, and the aim of the power should be guaranteeing possibilities of individual self-expression and self-realisation, as well as the value of the human being. The distortion of the mentioned aim can lead to the distortion of the essence of the discussed whole phenomenon, which, in turn, will make the necessity of its existence senseless.

The principle of separation and balance of powers is necessary not for the final isolation of the branches of power from each other, but for ensuring their interaction and cooperation. The main aim of this principle should not be the existence of mechanisms of mutual control, but first and foremost, finding more acceptable and effective joint solutions to problems.

Any crisis or an emergency situation is difficult, requires quick reactions. This very often leads to massive restrictions of human rights. It should be noted that such situations do not give an opportunity to restrict the fundamental rights, ignoring the criteria necessary for making the process legitimate, or violate other principles and values necessary for democratic ethics. The rules, which should be followed in those situations, where developed with the aim of eliminating violations in such processes. Hence, they should be strictly followed even in such an unprecedented situation, which we have been facing since last year.

The presented circumstances should be the axis for democratic ethics, and in case of their implementation, democracy will become a reality in the modern world.

THE DOCTRINE OF FRUSTRATION IN SPANISH LAW: ITS CONFIGURATION IN LIGHT OF THE PANDEMIC

*JUAN PABLO MURGA FERNANDEZ**

ABSTRACT

The global pandemic of 2020, and the measures taken to combat it, are generating a variety of issues in relation to the performance of contracts. This work analyses a mechanism which often "wakes up" in times of crisis, *viz* the doctrine of frustration, or *rebus sic stantibus clause,* under the lens of Spanish law. It analyses the basis of the clause, its configuration in Spanish law and reflects on its usefulness as an approach.

INTRODUCTORY PREMISES: PANDEMIC AND THE LAW OF CONTRACTS

The year 2020 has marked a watershed moment for the world's population, with an unparalleled public health crisis which has paralyzed and transformed a normality that we assumed to be immutable. The world of law is no stranger to this reality and its effects have been profound in relation to contract. The pandemic caused by COVID-19, alongside the measures taken to combat it (lockdowns, restrictions on freedom of movement, prohibition of commercial and social activities[1]) are raising many problems in relation to the performance of the contracts pending execution. What response does traditional Spanish contract law offer in such situations? There are several general mechanisms called to play a fundamental role in this context: *force majeure*, good faith and *the rebus sic stantibus clause* (also known as the doctrine of excessive onerousness or frustration).[2] We will focus our attention on this last institution, which in times of crisis usually awakens from the lethargy that ordinarily characterizes it.

It is a doctrine with diverse facets, and in the context of Spanish law has the particularity of not being recognized in the articulation of our Civil Code.[3] Indeed, the *rebus* is the creation of the case-law of the Supreme Court, dating from the inception of the Civil War. Unsurprisingly therefore, it has undergone significant developments over the years. Nonetheless, there is broad consensus that it has been effectively adopted into the norms

*Senior Lecturer of Civil Law, University of Seville. jpmurga@us.es

I would like to express my deep gratitude to the Profs. Javier Garcia Oliva of the University of Manchester and Helen Hall of Nottingham Trent University, for the invaluable help provided and the time in the preparation of this article. I would also wish to acknowledge this paper has been carried out within the framework of the National Research Project «Subjects and Instruments of Private Traffic VII: Real Estate Market and Economic Crisis (DER2015-66043-P)». On the subject addressed in this study, see our recent work "The -manida- clause rebus sic stantibus: reflections on its usefulness in times of pandemic", in Cerdeira Bravo de Mansilla G. Garcia Mayo M. Coronavirus y Derecho en Estado de Alarma, Reus, Madrid, 2020, 341–362.

[1] Gregoraci, B., "El impacto del COVID-19 en el Derecho de contratos español", *Anuario de Derecho Civil*, Volume LXXIII, Fasc. II, 2020, 457–458, summarizes the measures taken by the Government by Royal Decree 463/2020 of 14 March declaring the state of alarm for the management of the pandemic caused by COVID-19, which impact on private contracts: "Containment measures, adopted by the Government in the [. . .] Royal Decree 463/2020, which have impacted contracts have been the suspension or reduction of commercial activity . . . and the restriction on the freedom of movement of persons".

[2] Gregoraci, B., "El impacto del COVID-19 en el Derecho de contratos español" . . . , *Cit*., 458.

[3] Note, however, that Navarrese civil law has regulated the *rebus sic stantibus* Law 498 of the Compilation of Civil Foral Law of Navarra (introduced by the Law Foral 21/2019, of April 4, on amendment and updating of the Compilation of The Civil Law Foral of Navarra or Fuero Nuevo), which states: "In the case of long-term obligations or successive contracts, and during the time of fulfilment, the economic content of the obligation or proportionality between benefits is fundamentally and seriously altered, as unforeseen circumstances have rendered compliance by one party extraordinarily onerous, that party may request judicial review to amend the obligation in terms of fairness or declare its decision".

of the juridical framework. It has been extensively discussed by generations of scholars, and could never be accused of being a recent or uncertain innovation. Proof of its status can be found in the seventh additional provision of Law 3/2020 of 18 September, setting out procedural and organisational measures to deal with COVID-19 in the field of the Administration of Justice (*Ley 3/2020, de 18 de septiembre, de medidas procesales y organizativas para hacer frente al COVID-19 en el ámbito de la Administración de Justicia*). This states:

> The Government shall submit to the Justice Commissions of the Congress of Deputies and the Senate, within a period of not more than three months, an analysis and study on the possibilities and legal options, including those existing in comparative law, to incorporate into the legal regime of obligations and contracts the rule rebus sic stantibus. The study will include the best available data on the impact of the COVID-19 crisis on private contracts.

In the following pages we intend to focus on this doctrine in the light of the pandemic: we will analyze its origins, its configuration in Spanish law and its practical and fiscal implications of its application in the current crisis. Then in our conclusion will determine whether if it is a useful and appropriate cure for the infection which has permeated contracts.

THE REBUS *SIC STANTIBUS CLAUSE:* EXCEPTION TO THE MANDATORY FORCE OF THE CONTRACT AND ITS DEBATED BASIS

The starting point for the study of the *rebus clause is* the reference to the principle of binding or compulsory force of contracts, known under the Latin aphorism of the *pacta sunt servanda,* and essentially contained in art. 1091 CC: *"*The obligations a arisen from contracts have the force of law between the contracting parties, and must be fulfilled within the meaning of the contracts*"*.[4] This is a key principle not only for law, but also for the economy, since the security of economic transactions is closely linked to the trust that can be placed in the effectiveness of enforcement of contractual obligations. If it were easy to repeal the *pacta sunt servanda*, neither credit nor capital markets could operate: it is clear that a credit agreement whose repayment and interest with only doubtful enforceability would have little appeal for lenders.[5] The underlying importance of this principle explains its quasi non-derogable nature in the original formulation of the generality of the European [6] Decimononic Codes and also the Common

[4] Article 1091 CC is the most representative of the principle of binding contracts, although it is also reflected in other precepts of our Code: Art. 1278, 1256, 1258 and 1260. About this principle see the historical and comparative study carried out by Casas, M.G., *Der lukrative Schuldvertrag (Eine historisch-institutionelle Dekonstruktion seiner Physiognomie)*, Duncker & Humblot, Berlin, 2020.

[5] Salvador Coderch, P., "Alteración de circunstancias en el art. 1213 de la Propuesta de Modernización del Código Civil en materia de obligaciones y contratos", *Boletín del Ministerio de Justicia*, Year LXV, No. 2130, April 2011, 10; the same author stresses that "if all contracts were generally reviewable, the confidence of economic operators would fade; [. . .] trust . . . is fundamental in any economic system, as contractors are confident that contracts will normally be fulfilled without recourse to apparatus of the State, such as judges and courts . . . " (11).

[6] Note that we refer to the "original formulations" of the decimononic civil codes, since in the most representative codes the picture has changed substantially: this has been the case in France, following the very important reform of 2016, with the new art. 1195 of the *Code*; or Germany, which after the 2002 reform of its law of obligations also incorporates in § 313 BGB the *Störung der Geschäftsgrundlage* which had already been introduced by their case law. Zimmermann R. *The new German law of obligations. Historical and comparative perspectives* Oxford University Press, Oxford, 2005, 3: "There is on aspect of the reform legislation that should be mentioned [. . .] this is the incorporation intro the text of the BGB of a number of doctrines that had previously come to be recognized *Praeter Legem*; in particular: [. . .] change of circumstances (*Störung der Geschäftsgrundlage*. 313 BGB) [. . .]".

Law.[7] Exceptions to it were not permitted, required that a validly concluded contract be fulfilled in the agreed terms whatever eventualities might arise. The Spanish Civil Code follows the same pattern, clearly influenced by the French Code.

However, this fidelity to the principle of binding contracts is not boundless, and the other considerations may sometimes cause it to give way. Particularly where the circumstances in the minds of the parties at the formation of the contract, and the distribution of the risk agreed by the parties, have altered dramatically and beyond recognition.[8] In such cases, the obligations assumed may have transformed into something excessively onerous for one of the parties; or the purpose of contract may be completely frustrated. In other words, there is an essential alteration of the basis on which the contract was established. This is where the *rebus sic stantibus* clause, comes into play, which is referred to under various denominations at the comparative level: doctrine of excessive onerosity (Italy),[9] doctrine of frustration (Common law), doctrine of the modification of the basis of the legal business[10] (Germany),[11] or doctrine of unprevision (France)[12] among others.[13] In all cases, it is a matter of admitting an exception to the principle of the binding force of the contract, enabling the revision of the content of the contract and even its termination in the event of an extraordinary, supervening and beyond calculation change in the circumstances in which the contract was concluded and whose obligations have not yet been fully performed (war, economic crisis or public health disasters).[14]

One of the most discussed aspects of these legal figures, all of which respond to the same common denominator (i.e. a substantial change in contractual circumstances) is the determination of their basis**.** This is particularly relevant to systems such as the Spanish one where the *rebus clause is* not expressly recognized in a normative text.[15] There are several doctrinal constructions in this regard on which rivers of ink have been spilt. In good overview of perspectives can be achieved by following Díez-Picazo, who places theories into two main categories: subjective and objective. Subjective theories seek their support in an implicit or implied willingness of the contracting parties to

[7] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial*, Volume II (Las relaciones obligatorias), Civitas, Cizur-Menor, 2008, 6th Edition, 1057: "What is the influence that an alteration of circumstances must exert on the life of the contract or the obligatory relationship? This is a problem that laws and civil codes enacted during the 19th century, at time of some economic and social stability, did not address". Salvador Coderch, P., "Alteración de circunstancias en el art. 1213 de la Propuesta de Modernización del Código Civil en materia de obligaciones y contratos" . . . , *Cit*., 12: "If the historical optimal balance has ceased to be applicable, then clinging to the past may be counterproductive".

[8] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., 1057, indicates ". . . seems clear that, at least in certain cases, unconditional fidelity to the contract may lead to consequences that clearly appear to be unfair".

[9] The *"eccessiva onerosità soprravenuta"* regulated in articles 1467–1469 Italian Civil Code.

[10] On the treatment of the change of circumstances in English law., Cartwright J. *Contract Law: An Introduction to the English Law of Contract for the Civil Lawyer*, Hart Publishing, Oxford, 2016, 3rd Edition, 260–271.

[11] The *"Störung der Geschäftsgrundlage"* contemplated in § 313 BGB.

[12] The *"théorie de l'imprévision"*, Article 1195 of the *Code* (provision introduced with the reform of the law of obligations 2016). A detailed analysis of the reform can be seen in Cartwright, J. Whittaker, S. (Editors), *The Code Napoléon Rewritten. French Contract Law after the 2016 Reforms*. Bloomsbury, Oxford, 2017; about the novel theory of *imprèvision* which had already been recognised by French jurisprudence, particularly the chapter on Fauvarque-Cosson, B., "Does review on the ground of imprevision breach the principle of the binding force of contracts?", in Cartwright J. Whittaker, S. (Editors), *The Code Napoléon Rewritten. French Contract Law after the 2016 Reforms*. Bloomsbury, Oxford, 2017, 187–205.

[13] Parra Lucán, Mª. A., "Riesgo imprevisible y modificación de los contratos", *In Dret*, No 4, 2015, 4, stresses that "the terminological difference and translation difficulties are largely due to [. . .] to the lack of identity of the assumptions included in the various doctrinal, jurisprudential analyses and in the reference texts in the processes of harmonisation of contract law".

[14] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., p. 1056: [. . .] In times of serious disruption, (such as wars, calamities, economic crises), the problem becomes especially pressing and cannot be solved simply by the rule of unconditional fidelity to the contract; especially in cases where changes, political, economic, social or technological, alter the basis on which the initiation and maintenance of contractual relations had been established."

[15] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., 1068, stresses that "the problem of the technical and legal basis is particularly serious in those systems where the possibility of amending the regime established by the voluntary agreement is not found in legislation".

insert into the contract a clause under which the subsistence of the contractual relationship depends on the correlative subsistence of certain circumstances existing at the time of conclusion of the contract and the variation of which is not foreseeable. For their part, objective theories seek the basis of the *rebus* through two fundamental pathways: the principle of good faith and the supervening failure of the causal mechanisms of the transaction. From the perspective of good faith, a consequence of this principle which governs the entire life of the contract (from its negotiation to its final performance) is the restoration of the contractual basis with equitable reciprocity of obligations. If a *causal* approach is taken, the economic-social function of the contract comes to play, and it is argued that this essential element must apply not only at the time of formation, but also throughout the life of the contract. On this basis, the contract must end in the event that this foundational element crumbles.[16]

The objective theorists are undoubtedly in the majority, which is unsurprising in light of the inconsistency which besets the subjective analysis. As Díez-Picazo points out, it is difficult to follow the logic of purporting to found a doctrine on the fictional will of the parties on the occurrence of an event which is accepted to have been unforeseen.[17]

THE *REBUS CLAUSE IN* SPANISH CASE LAW: BRIEF NOTES ON ITS EVOLUTION AND THE BASIS FOR ITS IMPLEMENTATION

Bearing in mind the context in which the *rebus* arises and academic controversy over its theoretical foundation, which clearly reflects its character as applying an exception to the cornerstone principle of the binding force of contracts, we now turn to study its origins and current configuration in Spanish law. We must refer to the case-law of the First Chamber of the Supreme Court and the developments there, since as previously noted, this doctrine is rooted in case law.

Leaving aside the historical origins of the *rebus clause,* which go back to canon law, in Spanish law the leading case is "More c. Carsi" decided by the STS on June 30, 1948.[18] It was a case in the plaintiffs and defendants formed a contract, which included an option to purchase a plot of land. The plaintiffs, who alienated their adjoining property, nonetheless claimed compliance with the option to purchase. The application was dismissed on the basis that the interest of the plaintiffs in the acquisition was based on enabling an exit and facade to their adjoining property on a new street, so that having alienated that property, it was no longer possible to fulfil the purpose intended when the contract was concluded.[19]

From this moment on, three distinct phases can be distinguished in the case-law of our High Court: an initial one, which covers the period from the 1940s until the year 2013, during which *the rebus* clause is conceived in very exceptional and restrictive terms; a second stage, which comes to be classified as the period "normalization" for the doctrine, when it acquired some flexibility. This began in 2013, and lasted until the advent of the current phase, which we shall shortly discuss. Firstly however, we need to consider each component stage in more detail.

The first stage lays its foundations in the original decisions of the Supreme Court, which are given in the context of the situation of economic instability of the post-civil war period. In these rulings, the doctrine of the *rebus* clause is configured in a very

[16] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., 1068–1069.
[17] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., 1068.
[18] RJ 1948–1115.
[19] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., 1057.

restrictive and cautious way.[20] Evidence of this, can be found in the refusal of courts to recognize situations such as the weakening of the currency or the elevation of rail[21] fares as satisfying the threshold of unforeseeability[22] necessary to trigger the doctrine.[23] A good summary of the main characteristics of this phase of development can be seen in the ruling of STS of 17 May 1957:

> A) That the rebus *sic stantibus clause* is not recognized in legislation;
> B) That, however, given its development from case law and the principles of justice which serves, there is scope for it to be elaborated and accepted by the Courts;
> C) That this doctrine has a potentially dangerous application, and, even where its consideration is appropriate, caution must be exercised in concluding that it applies;
> D) That its acceptance requires as fundamental premises:
> (a) extraordinary alteration of the circumstances at the time of performance of the contract in relation to those concurrent at the time of its conclusion;
> (b) a gross distortion of the benefits and burdens between the parties, outside of their contemplation and rendering the new balance so disproportion as to undermine the principle of reciprocity ;
> (c) that all of the foregoing spring from extreme and unpredictable events;
> E) As regards its effects, it is aimed at compensating for the imbalance of benefits and burdens. [24]

From 2013 on, a new stage was beginning to be glimpsed in the jurisprudence of our High Court, which sought [25]to consolidate itself in particular with two 2014 rulings.[26] In the 2013 judgments, it is first acknowledged that the economic crisis can open up a scenario of recognition of amending effects for unforeseen extraordinary alterations. This gives rises to the possibility of considering the economic crisis as the basis for the application of the *rebus sic stantibus doctrine,* breaking out beyond the boundaries of the classic formulation and its requirement of extreme and exceptional circumstances. The following passage from the STS ruling of 17 January 2013 is typical of the kind of reasoning being advanced at this stage:

> An economic recession such as the current one, with profound and prolonged effects, can be classified, where the contract has been concluded before the crisis was widely perceived, as an extraordinary alteration of the circumstances, capable of triggering, provided that other requirements are met in each individual case [. . .] a gross alternation of the benefits and burdens between the performance of the parties, so as to render the same disproportionate. [27]

[20] Salvador Coderch, P., "Alteración de circunstancias en el art. 1213 de la Propuesta de Modernización del Código Civil en materia de obligaciones y contratos " . . . , *Cit*., 14, considers that "in Spain, the doctrine of the *rebus sic stantibus* it was set out in case law of the Supreme Court after the Civil War crisis (1936–39), but its conceptual reconstruction took place with rigour."

[21] STS of June 5, 1945 (RJ 1945–698).

[22] STS of May 17, 1941 (RJ 1941–632).

[23] Espín Alba, I. *Cláusula rebus sic stantibus e interpretación de los contratos: ¿y si viene otra crisis?*, Reus, Madrid, 2020, 102. The same author carries out a thorough analysis of the jurisprudential developments experienced by the clause (101–147). On this evolution of case law., see also the study by Marañón Astolfi, Mª., "Evolución doctrinal de la cláusula *rebus sic stantibus* en la jurisprudencia del tribunal supremo. Comentario a la Sentencia del TS de 6 de marzo de 2020 (JUR 2020/89493)", *Aranzadi Heritage Law Magazine* N º 52, 2020 (BIB 2020–34124).

[24] RJ 1957, 2164.

[25] The judgments initiating the anteroom of this new stage are the SSTS of 17 and 18 January 2013 (RJ 2013–1819; RJ 2013–1604, respectively); see Espín Alba, I. *Cláusula rebus sic stantibus e interpretación de los contratos: ¿y si viene otra crisis?* . . . , *Cit*., 110–111.

[26] These are the SSTS of 30 June 2014 (RJ 2014–3526) October 15, 2014 (RJ 2014–6129); rulings by the then Magistrate rapporteur, Francisco Javier Orduña Moreno, who had already anticipated his normalizing stance on the clause *Rebus* in previous doctrinal works; among others, it is worth noting: Orduña Moreno, F. J., Martínez Velencoso, L. M., *La moderna configuración de la claúsula "rebus sic stantibus": (tratamiento jurisprudencial y doctrinal de la figura)*, Civitas, Cizur-Minor, 2003.

[27] RJ 2013–1819.

Further judgments in 2014, consolidated a more flexible interpretation of extraordinary alteration of contractual circumstances. Espín Alba, postulates that this "normalizing" of the *rebus clause* can be summarized in the following points:

A) Recognition of the economic and financial crisis as a notorious fact that must be taken into account for the purposes of analyzing the circumstances that have been arisen.
B) The cases dealt with in the 2014 judgments were not subject to consumer law, meaning that the principle of private autonomy governs the determination of contractual content, in particular for clauses distributing risks or limiting the liability of one of the parties. For the undertakings concerned the subject-matter of the contract was part of their ordinary activity, and in some instances part of the very essence of the social object of the organisation.
C) This modern configuration of the *rebus sic stantibus clause* is linked to both objective and subjective theories, integrating good-faith and reciprocity into the construction of agreements. Moreover, the effects of the doctrine essentially have the effect of modifying agreements, rather than terminating them.[28]

In short, with this new position the Supreme Court aimed to transform the highly exceptional nature of the *rebus* clause*,* normalizing and making its application more flexible in construing which amounts to a dramatic alteration to the factual conditions under which the contract was concluded (allowing the economic crisis itself to act as a trigger for the *rebus).*

This normalization brought with it widespread claims based on *the rebus sic stantibus* clause, which stepped into the limelight as the main argument in a multiplicity of cases on disparate matters, and attempts were made to use it as a panacea for all the of the ills arising from the economic crisis (e.g. denial of payment without adjustment of fees, interest, or review of annuity).

This new judicially innovated doctrine, however, was not boundless in scope. As far back in 2015 and the even more firmly in 2019 and 2020, the High Court steered back towards a more the restrictive path, reconfiguring the *rebus* clause once more into an exceptional mechanism applied only in extremis. The first exponent of the progressive abandonment of the normalizing trend comes from the STS ruling of 24 February 2015.[29] The judgment concerned the sale of real property in the countryside for the purposes of urban development, with the seller demanding that a stalling buying should fulfil their obligations. The defendant buyer argued for a 50% reduction in the price under the *rebus sic stantibus clause* on the basis of alteration of the circumstances following the bursting of the housing bubble. The Supreme Court dismissed the buyer's claim, on the basis that fluctuations in the housing market constitute a normal risk inherent in such contracts, particularly where the purchaser is a company whose activity is part of the real estate sector. There are several passages of that judgment which flag up the conscious abandonment of a more routine application of *the rebus sic stantibus clause.*

It is noted that even though the development of case-law had trended towards "a more flexible characterisation and appropriate to its nature, its prudent and moderate application remained key, given the requirements of its specific and delineated basis, as well as the necessity to keep it within concrete bounds in light of its impact upon the enforceability and certainty of contracts" has also been retained. It is added, moreover, that "the consideration by that Chamber of the notorious *de facto* nature which characterized the economic crisis of 2008 does not, on its own, entail a generalised, or automatic, application of the *rebus sic stantibus* clause from that period. In reality it

[28] Espín Alba, I. *Cláusula rebus sic stantibus e interpretación de los contratos: ¿y si viene otra crisis?. . . , Cit*., 114.
[29] RJ 2015–1409; the judgment was, curiously enough, drafted by the same Orduña Moreno.

is vital that its causal relationship or actual impact in the context of the contractual relationship in question, can be outlined."[30] This retrenching trend strengthened even further in recent judgments of 2019 and 2020.[31]

It is worth noting the STS of 6 March 2020, the last decision[32] in which our Supreme Court has ruled on the *rebus clause* (curiously only a few weeks before the outbreak of the pandemic caused by COVID-19). The judgment concerned the interpretation of contracts on the exclusive transfer of the management, promotion and sale of advertising spaces, for the issuance of advertisements on television and radio channels owned by the principals. Pursuant to the agreements, television and radio channels undertook to pay a percentage and escalation commission on the gross revenue from the advertising. At the same time, the applicant guaranteed a minimum production consisting of the acquisition of advertising spaces for a "guaranteed minimum" of the gross amount.

The "guaranteed minimum" for each year increased or decreased in direct proportion to the variation of audience figures or average annual screen share over two years. After concluding the contract, the annuities were provided the years 2006 and 2007. The then applicant continued to provide its services for the two public bodies, and invoiced them at the same percentage rates as for the 2006 contracts, but the parties failed to agree on the guaranteed minimum for the 2008 annuity.

In 2008 there was a decrease in advertising investment compared to the previous year. Consequently, the applicant brought an action against public television and radio authorities for recovery of the amount of the outstanding invoices for the services provided in the procurement of advertising. The defendants opposed the application and issued a counterclaim. The core dispute revolved around the fixing and determination of the "guaranteed minimum" to be understood in force. The judgment of first instance required the broadcasters to pay the amount of the outstanding invoices to the applicant but at the same time also partly upheld the counterclaim, and ordered the advertising company to pay an amount reflecting the net reduction in income for 2008.

The Appeal Court partially allowed the appeal, arguing that while the applicant advertising company had to pay the radio-television broadcaster an amount corresponding to the "minimum" set, there had nevertheless been a breakdown in the economic basis of the contract, resulting from the "*fall in the advertising market*". This meant that the risk assigned by the contract to the advertising entity "must not completely destroy the reciprocity of the contract to the point of completely neutralizing the expectation of profit of one of the parties and even hold it accountable to the principal for the total decline in advertising revenue caused by external circumstances that are completely beyond its control".

The Supreme Court then partially reversed this finding, denying the application of the *rebus* clause in the case at issue. The reasoning used by the TS to reach such a conclusion is condensed into the following passage:

> According to the caselaw on the *rebus sic stantibus* clause, an alteration of the circumstances leading to the amendment or, ultimately, the termination of a contract, must be of such a magnitude that it significantly increases the risk of frustration of the purpose of the contract. Of course, such circumstances must be totally out with the contemplation of the parties [. . .] . The unpredictability of the change of circumstances is necessary for the

[30] Passages highlighted by Espín Alba, I. *Cláusula rebus sic stantibus e interpretación de los contratos: ¿y si viene otra crisis? . . . , Cit*., 121.

[31] SSTS of 18 July 2019 (RJ 2019–3010) March 6, 2020 (RJ 2020–879).

[32] RJ 2020–879.

> application of the "rebus" rule. If the parties have expressly or implicitly assumed the risk that a circumstance would have happened or should have assumed it because, under the circumstances and/or nature of the contract, such a risk was reasonably foreseeable, it is not possible to assess the alteration that, by definition, implies the non-assumption of the risk [. . .]. There can be no mention of unforeseen alteration when it is within the normal risks of the contract [. . .]. The change in those characteristics which, under the premisses established by the case-law, could lead to a case of application of the rule of the *rebus sic stantibus* is more likely to take place in a long-term contract. But not in a case, such as the present one, of a short-term contract, in which something extraordinary affecting the basis of the contract has less scope to arise. In our case, where the duration of the contract is one year, since it is the annual extension of an initial contract lasting two years, it is difficult for a change in circumstances relating to the demand on the market for TV advertising, to be construed as being outside of the risk assumed by the extension of the contract. [. . .] The decline in demand for TV advertising, coming forward for a short period of time, a year, was a risk covered by the contract. Furthermore, the eventuality which materialized was neither drastic nor unpredictable: the decrease in advertising investment in general was 25.9 million euros in 2007 to 24.1 million euros in 2008. Consequently, the rule rebus sic stantibus does not apply, which is why we allow this appeal [. . .].[33]

As can be seen, the High Court emphasises the distribution of the risks associated with the contract: if, in the light of the circumstances, it is determined that the risk has been expressly or implicitly assumed by one of the contracting parties, the possible application of the *rebus* clause *shall be excluded.* This aspect seems to obviously greatly reduce the practical operation of the *rebus* clause. In light of this decision, when will it be understood that an overselled and extraordinary event exceeds the threshold of the risks that the parties distribute in exercise of the their autonomous free will when concluded the contract? We will see later whether the pandemic and the measures taken to combat them exceed that threshold.

This ruling also clarifies nature of the contracts amenable to amendment via the application of the *rebus* clause*:* they must be long-term contract (successive contracts, or a single contract with deferred execution). The latter point may have to be revisited in light of the pandemic context. The bottom line is also that the classic approach of the Supreme Court is more strongly reasserted than ever.

Having concluded this brief analysis of the developments in the case-law on the *rebus sic stantibus* clause, it is worth pausing to address the general requirements required to implement it. These are conditions that have remained more or less uniform over the years, regardless of whether their setting varies substantially in view of the configuration (exceptional or standardized) of the *rebus.* Díez-Picazo, observes the following:

a) The obligatory relationship envisaged must be of a long-term nature, i.e. a lasting compulsory relationship.
b) The obligatory relationship in question must be pending performance in whole or in part. The disappearance of the basis of the transaction shall affect the obligations still to be performed, but not the obligations already executed.
c) There must be a supervening disappearance of the basis of the transaction, which occurs when the following circumstances are met:
    i. Where the *ratio* of equivalence or proportion between benefits is totally destroyed in such a way that the benefit and consideration cannot be properly discussed. This is identified by the so-called "excessive onerosity" of the contract, which may result in the increase in the cost of the benefit for one of the parties, or in the gross reduction in its value to the other.

[33] Legal Basis Fourth, point 2.

ii. When the common purpose of the contract, expressed in it, or the substantial purpose of the contract for one of the parties, which is admitted and not rejected by the other, is unattainable. This is where the objective basis of the *rebus* clause comes into play, which either on the basis of the principle of good faith or of the cause that presides over the whole life of the contract, is identified with a breach of the purpose of the contract (by mutual agreement, or for one party with the non-opposition of the other party).

d) This alteration of the basis of the contract must be the result of an extraordinary change in the circumstances, in relation to those existing at the time of the conclusion of the contract; and, furthermore, it must be radically unforeseen and unforeseeable at the time of its occurrence. Consequently, it excludes circumstances that were foreseeable, or that were in the sphere of control of the injured party. Changes in circumstances should also be disregarded where the risk constitutes the determining reason for the business relationship: this is the case in one off, sole performance contracts or where the alteration in question forms part of the risk assumed by one of the parties according to the nature of the type of contract.[34]

Let us then consider whether these criteria are fit for purpose in the current context, in view of the effects of the pandemic; and value the advantages of the possible positivization of *the rebus sic stantibus clause* in the articulation of the Spanish civil code.

THE *REBUS CLAUSE* BEFORE THE COVID-19 PANDEMIC IN SPANISH LAW: BALANCE SHEET ON ITS USEFULNESS AND THE POSSIBLE NEED FOR ITS REGULATION

As we indicated at the beginning of our presentation, *the rebus sic stantibus* clause is a figure that is reborn in times of crisis. The current pandemic is no exception in this regard, hence it is one of the institutions that has generated the greatest doctrinal debate in Spanish contract law following COVID-19 tragedy.[35] Two essential aspects are at issue: Spanish civil code and, on the other hand, the usefulness of the *rebus clause* in dealing with the damage caused to those contracts.

Before answering these questions, three important clarifications should be made: firstly, we must keep in mind that what caused the arguably extraordinary alteration of the circumstances of the contract was not the pandemic itself, but the measures that have been taken to combat it.[36] Moreover the general answers we offer when analysing the assumptions of rebus in the light of the pandemic can only be considered as accurate through a case-by-case analysis of the different concrete cases in which the *rebus* clause is invoked (this is a matter where the solutions only serve as a general guideline, which must therefore be specified according to the circumstances of the case). Finally, it is important to define the scope of action of the *rebus* clause in respect of other mechanism arising in similar paradigms: such is the case of *force majeure*[37] which, unlike *the rebus,*

[34] Díez-Picazo, L. *Fundamentos del Derecho Civil Patrimonial* . . . , *Cit*., 1069–1070.

[35] See Gregoraci, B., "El impacto del COVID-19 en el Derecho de contratos español" . . . , *Cit*., 463.

[36] See Alfaro, J. and Garicano, L., "El legislador bondadoso en la epidemia", *Almacén de Derecho*, entry of April 3, 2020: "The change in circumstances is not the epidemic. The change in circumstances is the government order issued under the alarm statement to close the establishments." See also Gregoraci, B., "El impacto del COVID-19 en el Derecho de contratos español" . . . , *Cit*., 460, in relation to force majeure in times of pandemic: "It is therefore possible to state that what in each contract may constitute force majeure will be the concrete measures taken as a result of the health emergency arising from the pandemic.

[37] In addition to the *rebus* clause it has been the other major institution of contract law discussed at length in the context of the pandemic. On this subject, see Carrasco Perera, A., "Permítame que le cuenta la verdad sobre covid-19 y fuerza mayor", Cesco, April 17, 2020, who begins his practical analysis in the following terms: "If it were not for the fact that the bars are closed, even the parishioners would talk about the Force Majeure. It has gone viral. Experts and heathens are pontificating at the end of the COVID-19. The noise is making a mess of things."

renders performance completely impossible (with the rebus, the extraordinary change of circumstances leads to a breach of the basis of the transaction for the reasons already analysed, although the performance itself remains enforceable).[38]

Considering again the requirements of the *rebus* clause in connection with the contracts affected by the pandemic it can be anticipated that many of the cases may be covered by it (at least abstractly, at the expense of analysing whether the measures adopted to deal with the pandemic exceed the threshold of the risk assumed by the parties to the contract). We should therefore justify our position by some nuanced discussion of the topic of the *rebus sic stantibus* clause in the current circumstances:

— There are numerous contracts that were concluded prior to the declaration of the alarm status and were still pending execution at the watershed moment.
— In order for the measures taken by the State administration (which took sole command when the alarm status was decreed) to alter the circumstances of conclusion of the contract in an extraordinary way, it is not necessary for the contract to be of "long duration"; it is sufficient that the contract is pending performance, regardless of whether it is of short duration.[39] Within days, a myriad of contracts were severely affected. Hence, the duration of the agreement should not be decisive in contexts of sudden and unpredictable calamities such as those of the current pandemic. It is sufficient that in the time between perfection and execution an event takes place that substantially alters the contractual circumstances (that is precisely what has happened with the pandemic).
— The pandemic and the measures taken following the declaration of the state of alarm have in many cases caused an alteration of the circumstances that can be considered extraordinary, overriding and completely unpredictable (Could anyone, even in January 2020, have ever contemplated the situation which has endured since March? Aside from perhaps a few experts in the field of infectious disease and disaster planning, nobody could have predicted the events which unfolded).[40] Moreover, this alteration has destroyed the basis of many

[38] Parra Lucán Mª. A., "Riesgo imprevisible y modificación de los contratos" . . . , *Cit*., 21, highlights the differences between the clause *rebus sic stantibus* and force majeure, and highlights the difficulties presented in many cases by such a distinction: "The requirement of unpredictability raises the need to define the scope of the circumstances that allow the contract to be amended in the fortuitous case (or force majeure) which, in Spanish law, unless he had contractually assumed that risk (for example, by charging a premium for the difficulty of the circumstances), releases the debtor who can no longer comply (arts. 1182 and 1184 CC) and exonerates him liability for damages suffered by its creditor (art. 1105 CC), unless temporary and then only suspends compliance. The difference between the type of problem referred to in the *Rebus* and the fortuitous case has more to do with the outcome (impossibility of compliance in the fortuitous case) than with the reason that causes it. But the scope of the fortuitous case and that of the *Rebus* can be difficult to de-armor. This is for two reasons: on the one hand, because the line separating impossibility from extraordinary difficulty is sometimes very thin; another, because both categories share the budget of unpredictability or inevitability."

[39] See. Martinez by Aguirre Aldaz, C., "Modificación del contrato", in Martinez by Aguirre Aldaz, C. (Coordinator), *Curso de Derecho Civil (II). Derecho de obligaciones*, Colex, Madrid, 2014, 4th Edition, 472, echoing the position also held by Prof. Albaladejo: "Although this will be the norm, these do not need to be long-term obligations. This is a requirement commonly contemplated by doctrine . . . and jurisprudence [. . .]. In the face of this, Albaladejo has argued that they do not need to be contracts of continuous execution, but that it is sufficient that it is between its conclusion and compliance a time when altering circumstances is possible."

[40] See Rodriguez Rosado, B. and Ruiz Arranz, A. "Consecuencias de la epidemia: reequilibrio contractual y Covid-19", *Almacén del Derecho*, entry of April 16, 2020, defend the totally unpredictable nature of the COVID-19 pandemic: "It should be assumed that the risks arising from the coronavirus are not assigned in contracts that were consigned prior to the outbreak of the crisis: contractors were able to think and take other risks, risks, let us say so, minor and individual, such as illnesses, impossibility arising from accidents of chance or guilty, and even unexpected economic or financial setbacks (think speculative cutting contracts , such as swap activity) but not that economic and social activity is paralyzed entirely! It should also be noted that not even a war experience, however unlikely it may seem, would produce the economic effects of this universal confinement, since a war does not constitute an economic paralysis, moreover, in the short term, as Keynes knew, rather stimulates activity. The paralysis experience now produced and has not been previous, nor was it thoughtable. In this sense, to be purely contractually fulfilled (*lex contractus: dura lex, sed lex*) is absurd: the parties simply did not agree or foresee this."

> legal agreements: both by producing an unjustified imbalance of the benefits assumed by the parties (imagine the fixed monthly price paid by a hotel to a laundry for bed-linen . . . can the maintenance of that price be justly sustained with the drastic decline in the volume of tourists?). Equally, the underlying purpose of many contracts has crumbled. Think of the business premises lease contracts that were concluded just before the pandemic for the development of activities that are difficult to carry out – not impossible to carry out – is it possible to request the termination of the contract if the purpose of the contract, particularly on the part of the lessee, disappears? It will be necessary to analyse each contract in particular to give an exact answer; although theoretically, it should be reiterated that the *rebus sic stantibus* clause is perfectly applicable.

We emphasize the extremely fact sensitive nature of matter, because indeed, exact answers cannot be pointed out in a pressing way and in absolute terms. This is precisely what is highlighted by the authors when, for example, it is stated that "the possibility of some contracts considering risk allocation clauses drafted in very broad terms should not be ruled out, giving entry to any type of event (including COVID-19)". The key will lie, as is often customary in contractual disputes, in the careful interpretation to be carried out of the terms of each particular contract. However, this does not close the door in advance to the application of the *rebus clause;* rather, the essence of this mechanism (an extraordinary and totally unpredictable alteration of the circumstances that scuttling the basis of the agreement) must indisputably be a possibility with the advent of the pandemic.[41]

Thus, we come to the nub of our investigation: the *usefulness of the rebus* clause to "cure contracts infected by COVID-19"; which requires reference to its effects and practical implementation. If we come to affirm the admission of the *rebus sic stantibus clause,* what would its effective application look like?

The futility of the *rebus* clause in times of pandemic is revealed when we consider the possibility of judicial decision-making and the termination of contracts : we would make the healing of the damage suffered by the innocent contracting parties subject to the to a decision on the alteration of circumstances where the outcome will be uncertain, late and consequently totally unsatisfactory. In this sense, we follow the arguments put forward by Morales Moreno on the inadequacy of the *rebus clause* in dealing with the holes which that the pandemic has torn in the contractual fabric:

a) First, because it is inspired by commutative justice criteria (it is designed for each contract and not for the whole contractual network).
b) Second, because its application is slow (a reality exacerbated by an inevitable avalanche of claims), expensive and can lead to a litigation lottery, in light of the subjectivity of judicial decision-making and unpredictability of outcomes.
c) Third, because the application of *the rebus* doctrine does not allow for a metaphorical pause button to be pressed in relation to contractual performance, and this would be an invaluable device to have available. Even at the most critical moment of the public

[41] Morales Moreno, A.M., "El efecto de la pandemia en los contratos: ¿es el derecho ordinario de contratos la solución?", *Anuario de Derecho Civil*, LXXIII, 2020, fasc. II, Q. 450, indicates that the *rebus sic stantibus* it does not serve to solve all the problems that the pandemic causes in contracts pending execution; in particular, it denies its virtuality in resolving the liquidity or vulnerability problems of the contracting parties: "Undoubtedly the change of circumstances of the present moment meets the requirements of unforeseeability and the severity of the *Rebus*. But this does not exhaust all the requirements of its application. The proper case of application of the *rebus* (without prejudice to other uses of it) is the alteration of the equilibrium of the contract's performance, established at the time of its conclusion (the alteration of the basis of the transaction or the assumptions made). On the other hand, it is not the alteration of the contractor's economic capacity or liquidity to comply with, or his vulnerability. The correction of this important consequence of the crisis caused by the pandemic is therefore outside the scope of the *Rebus*, despite its importance. The *Rebus* (applied with legal rigour) does not serve to solve all the problems posed in contracts by the alteration caused by the pandemic".

> health crisis a party would not be justified in suspending performance. The parties remain obliged to fulfil the contract until the procedure invoking the *rebus* clause is completed.[42]

Given our discussion on the appropriateness of the *rebus* in times of pandemic, does it make sense for the clause to be incorporated the Spanish Civil Code? This is the process just launched through the seventh additional provision of Law 3/2020 of 18 September on procedural and organisational measures to deal with COVID-19 in the field of the Administration of Justice. The legal recognition of the *rebus clause* could therefore plausibly achieve a more formalised placed in the Spanish Legal Order. Most of the neighbouring juridical systems deal with the alteration of circumstances, as do important texts of European Soft Law.[43] The positive aspects of the standardisation of the clause would be as follows:

a) It would have a legal foundation, its articulation not reliant solely on fairness or contractual good faith. In addition, the endless discussions about the basis (objective *versus* subjective) of the clause would cease.
b) As a result, it would no longer be a dangerous, exceptional mechanism; a last resort of a subsidiary nature.[44]
c) The distance between the general rule (*pacta sunt servanda*) and the exception *(rebus sic stantibus clause)* would be reduced.[45]

However, this does not mean that the best time to legislate on the subject is in this era of uncertainty and great political, economic and social instability. This applies *a fortiori* when there is an absolute doctrinal consensus on the futility of the *rebus clause* to address the problems suffered by contracts during this pandemic. On the other hand, there is a precedent which could serve as a guideline for future reform, easing the process: Art. 1213 of the Spanish Proposal to Modernize the Civil Code on Obligations and Contracts. Even though this did not come to fruition, it received broad support and had sound doctrinal basis.[46]

Which leaves us with the riddle: what is the most appropriate remedy for the healing of contracts infected by COVID-19? Our solution is as follows: the intervention of

[42] Morales Moreno, A.M., "El efecto de la pandemia en los contratos: ¿es el derecho ordinario de contratos la solución?" . . . , *Cit*., 450–452. In the analogous sense, Ganuza, J.J., Gomez Pomar, F., "The instruments to intervene in COVID-19 time contracts: A Guide to Use", *In Dret*, N º 2, 2020, 561: "That is why we believe, with all sympathy, that those who undoubtedly laden with good intentions, confess to wishing to entrust the present task to general instruments, of imprecise profiles and in need of a singular decision adapted to the circumstances of each particular contract. In other words, the clause *Rebus*. This is not, we think, a useful mechanism in the current emergency. It's not because of its lack of Immediacy because of the unpredictable outcome and the overall cost to the judicial system and society as a whole, which involves subjecting hundreds of thousands, perhaps millions, of contracts to an uncertain and distant judicial decision on the matter. It does not even allow, in most cases, to narrow sufficiently clear negotiating ground to make it easier for the parties to reach a mutually satisfactory agreement to readjust the contract, except in a scenario such as the current scenario of alarming uncertainty about the duration and extent of disruptions to existing contracts. In short, the *Rebus* it is neither feasible nor convenient to address systemic events and macro shocks that suddenly impact an economy. Even less those phenomena that affect many (almost all, actually) contracts, but not symmetrically, uniformly and easily quantifiably but with hugely uncertain medium-term impact".

[43] Cfr. PICC (Article 6.2.2) PECL (Article 6:111), DCFR (Article III-1:110) and CESL (art. 89).

[44] Cerdeira Bravo de Mansilla, G. *Aspectos jurídicos del coronavirus*, Reus, Madrid, 2020, 195.

[45] Salvador Coderch, P., "Alteración de circunstancias en el art. 1213 de la Propuesta de Modernización del Código Civil en materia de obligaciones y contratos" . . . , *Cit*., 6–7.

[46] In recent times there are two proposals *lege Ferenda* on the possible normative formulation of the rebus clause: a) The first proposal of the Fide Crisis Office in civil and commercial matters on the legal regulation of the excessive onerousness of the performance and the frustration of the purpose of the contract in the light of the case law on the "*rebus sic stantibus*" clause (analyzed in detail by the Carrasco Perera, A., "Al fin la madre de todas las batallas del COVID 19: "Rebus sic stantibus". Con ocasión de una reciente propuesta institucional", *Cesco*, May 22, 2020. b) Principle 13.2 of the Principles developed by the *European Law Institute* COVID-19 crisis. It should also be noted that the regulation of the "rebus" was proposed July 2020 By Parliamentary Groups Citizens PdeCat and Compromís, who contributed a report prepared by Prof. Orduña Moreno (rapporteur of the controversial judgments trying to normalise the application of the clause *Rebus*), as a reference for future bills.

the legislature in order to allocate, under uniform criteria of distributive justice, the contractual risks of the pandemic.[47]

[47] Morales Moreno, A. M., "El efecto de la pandemia en los contratos: ¿es el derecho ordinario de contratos la solución?"..., *Cit*., 452. In this context, casting countless interpretative doubts, the necessary legislative measures that have been adopted in the field of private law are framed: suspension of eviction procedures; extraordinary contracts; automatic interest-free moratoriums; suspension of limitation and expiry periods, etc. A thorough and detailed analysis of the measures taken on contract law can be viewed in García Rubio, Mª. Q., "Medidas regladas en materia de contratos con motivo del covid-19 en España", *Revista de Derecho Civil*, No. 2, Special (May 2020), 15–46.

RES JUDICATA: A GAP BETWEEN CIVIL LITIGATION AND ARBITRATION?

*JESUS EZURMENDIA**

ABSTRACT

In the current English dispute litigation framework, arbitration has become undoubtedly the most utilised alternative to formal civil justice. Notwithstanding the well-recognised similarity between arbitration and litigation, some aspects of this analogy remain scantily discussed, particularly from the perspective of the final effects of the arbitral awards and how they could mirror civil courts´ finality and *res judicata*. The Arbitration Act 1996 does not expressly regulate the issue, laying the ground for the court to determine the applicability of preclusive effects to arbitral decisions. In this task, several doctrines have been examined, yet none of them have provided a conclusive response consistent with the overriding objective of civil justice and the broad relationship between both dispute resolution systems, particularly the assistance required from the court by arbitrators in the fulfilment of their assignment. This work attempts to critically assess the three main doctrines found in the English literature and legal authority, stating that the applicability of *res judicata* to arbitral awards should be based on the overriding objective stated in CPR 1.1 due to the public function of finality, rather than on purely arbitral grounds originating from a private agreement between the parties.

INTRODUCTION

This paper examines the applicability of *res judicata* to arbitration, specifically to arbitral awards. The work explores the current state of the discussion on these two key concepts of the civil dispute resolution framework today. By discussing the main positions that have tried to explain it and looking at selected critiques, this article aims to contribute to the debate and to a deepening of the analysis of finality in private systems of dispute resolution.

The relevance of the topic of study derives from the importance that arbitration and *res judicata* have in the legal system. Arbitration is now positioned as the most used alternative to court adjudication,[1] with the English legal system shifting away from its original reluctant approach to arbitration to a position of favouring it, broadening arbitrability.[2] For its part, *res judicata* provides the critical effect of finality, a consequence of a court's decision that ends the dispute for the parties and the state once and for all.[3]

Hence, *res judicata* is seen as a synonym for the conclusiveness of the judgment.[4] Accordingly, the habitual comparison of arbitration and litigation forces the question of whether arbitration is capable of bestowing similar effects to its decisions, in an

*Prof. Dr. Jesus Ezurmendia A. Bch. University of Chile, LLM, University College of London. Lecturer and faculty member at Universidad de Chile Law Faculty. Prof. Jesus Ezurmendia A. Lecturer at Procedural Law Department, University of Chile.

My acknowledgements go to Prof. Dr John Sorabji for his guidance, and Prof. Paula Nuno for her support throughout the process of research and writing of this article.

[1] Neil Andrews, *The Three Paths of Justice: Ius gentium* (Springer 2012) 220.

[2] David St John Sutton, Russell on Arbitration (24th edn, Sweet and Maxwell 2015) 72.

[3] K. R Handley, *Spencer Bower and Handley: Res Judicata* (Lexis Nexis 2009) 1.

[4] Rex R. Perschbacher, 'Rethinking Collateral Estoppel: Limiting the Preclusive Effect of Administrative Determinations in Judicial Proceedings' [1983] 35(3) UFLR 445–446.

effort to equate the efficacy of this private device to litigation. Thus, it appears to be necessary to examine the possible answers to this question and the rationales that could be given to explain it.

The focus of this work is the regulation of the Arbitration Act 1996 – hereinafter AA1996 – and the traditional twofold distinction of *res judicata* in England and Wales, as well as in most of the common law tradition, which divides it into cause of action estoppel and issue estoppel. Due to the extension and the main target of this essay, the newest branch of abuse of process is not specially considered.[5] Part I of this article briefly examines the origins, structure, the foundations, and extension of the effects of *res judicata*. This part also explains the institutions gathered under a broad concept of *res judicata* and how they work in the justice system. Part II concisely explains the main features of arbitration in perspective with civil adjudication, delving into their more significant differences as a way of presenting a clear background for the arguments provided subsequently in the essay. Part III is the core of the work. This part presents the relatively simple approach of English legal culture to the subject, mostly based on the *paradigm of similarity* between litigation and arbitration. The second portion of this section analyses and critically comments on the theories and doctrines founded in English statute and authority, as well as in domestic and international literature. Additionally, particular attention is given to issue estoppel and the consequences of its application. Part IV contains the author's opinion as well as a proposed rationale for the applicability of *res judicata* to arbitration. It lays the ground for a better understanding focusing on the strong relationship existing between arbitration and civil litigation and, consequently, how the aims and objectives of the latter have an impact on the effects of the former.

PART I – GENERAL BACKGROUND OF *RES JUDICATA*

Conceptualization and scope of res judicata

Endless litigation leads to chaos.[6] Every relatively sophisticated legal system in the world possesses a procedural tool that prohibits the relitigation of disputes that have already been settled by a valid court judgment. That institution is *res judicata*.

*Res judicat*a means, literally, 'a matter that has been adjudicated';[7] that is, once 'the *res* – thing actually or directly in dispute – has already been adjudicated upon . . . by a competent court, it cannot be litigated again'.[8] This is to say that judgements lead to an estoppel *per rem judicatum*.[9] *Res judicata* has been more specifically defined as 'the effect pronounced by a judicial or other tribunal with jurisdiction over the cause of action and the parties, which disposes once and for all of the fundamental matters decided (. . .) they cannot be re-litigated between persons bound by the judgement'.[10] Other formulations have asserted that: 'when a court of competent jurisdiction has determined, on its merits, a litigated cause, the judgement entered is forever and under all circumstances, final and conclusive as between the parties to the suit and their privies'.[11]

[5] The main case law of the abuse of process doctrine can be found in *Henderson v Henderson (1843) 3 Hare 100, 67 ER 313* and in *Taylor v. Lawrence [2002] EWCA Civ 90.*

[6] Allan Vestal, 'Preclusion/Res Judicata Variables: Adjudicating Bodies' [1966] 54(3) Georgetown Law Journal 857.

[7] Yuval Sinai, 'Reconsidering Res Judicata: A Comparative Perspective' [2011] 21(2) Duke Journal of Comparative & International Law 353.

[8] Adrian Zuckerman, *Zuckerman on Civil Procedure Principles of Practice* (Sweet and Maxwell 2013) 1238.

[9] *Ibid.*

[10] K. R Handley (n3).

[11] Robert von Moschzisker, 'Res Judicata' [1929] 38(3) Yale Law Journal 300.

Cause of action estoppel and issue estoppel

The wide concept of *res judicata* encompasses more than a single institution. Under its preclusive umbrella, English law comprehends a twofold division between cause of action estoppel in on the hand and issue estoppel on the other.[12]

The House of Lords provided a clarifying definition in *Arnold v. National Westminster Bank*. Lord Keith held that the cause of action estoppel:

> arises where the cause of action in the later proceedings is identical to that in the earlier proceedings, the latter having been between the same parties or their privies and having involved the same subject matter . . . [The] bar is absolute in relation to all points decided unless fraud or collusion is alleged[13]

The latter, issue estoppel, was defined in the words of Lord Keith in the same case as:

> Where a particular issue forming a necessary ingredient in a cause of action has been litigated and decided and in subsequent proceedings between the same parties involving a different cause of action to which the same issue is relevant one of the parties seeks to re-open that issue.[14]

And according to Lord Denning in *Fidelitas*:

> The law, as I understand it, is this: if one party brings an action against another for a particular cause and judgment is given upon it, there is a strict rule of law that he cannot bring another action against the same party for the same cause. Transit in rem judicatam. But within one cause of action, there may be several issues raised which are necessary for the determination of the whole case. The rule then is that, once an issue has been raised and distinctly determined between the parties, then, as a general rule, neither party can be allowed to fight that issue all over again. The same issue cannot be raised by either of them again in the same or subsequent proceedings except in special circumstances.[15]

As Zuckerman explains, parties are also bound by the findings of particular issues of law or fact considered essential to the final resolution contained in the judgment.[16]

THE PUBLIC FUNCTION OF *RES JUDICATA*

The doctrine of *res judicata*, as it occurs in several other aspects of the civil justice system, finds its underpinning rationale in a dual justification conveying both a private and a public perspective.

Private Justification

The private justification tends to safeguard the individual interests of the parties involved in litigation. The estoppel derived from *res judicata* signifies to the parties' protection from the annoyance, vexation, time and cost of repeated litigation.[17] As Vestal suggests, it 'saves litigants from multiplicity of litigation',[18] avoiding the disturbance of

[12] The terms 'claim preclusion' and 'issue preclusion' or 'collateral estoppel' could be used throughout this work, they are the terminology in American and international sources to refer to cause of action estoppel and issue estoppel respectively.

[13] *Arnold v. National Westminster* Bank, [1991] 2 A.C. 93 (H.L.) at [104–05].

[14] *Arnold v. National Westminster Bank*, [1991] 2 A.C. 93 (H.L.) at [105].

[15] *Fidelitas Shipping Co. Ltd v. V/O Exportchleb [1966] 1 Q.B. 630.* At 640.

[16] Adrian Zuckerman (n8) 1239.

[17] Robert von Moschzisker (n11) 299.

[18] Allan Vestal. (n6) 858.

defendants who have prevailed over dissatisfied litigants.[19] This idea is expressed as the Latin maxim '*debt bis vexari pro una et aedem causa',* which means that a litigant that has been through adjudicative proceedings has the right to not be bothered again for the same reasons. In Zuckerman's words, 'after judgment, parties are entitled to peace'.[20]

This approach is not only found in old Roman sources, but in common law it was recognised in 1599 by Lord Coke in *Ferrer v. Arden*, who conceived it as a guarantee to the right of the individual for protection against repetitive litigation. Protection is particularly necessary when asymmetrical litigants were party to the process. Thus, *res judicata* works as a shield against 'a rich and malicious man (who) would infinitely vex him (the defendant) who hath right by suits and actions'.[21]

Furthermore, the finality of *res judicata* allows the parties and their lawyers to correctly plan litigation.[22] When parties are aware that the judgment will bring a final and conclusive decision on the matter, they can properly choose their strategies more thoroughly, which may include the selection of an specific remedy, as it occurs in contract and tort cases, and to foresee the best possibility of enforcement.

Public Justification

From the public perspective, *res judicata* fulfils a critical function and is even more important than its private counterpart. The public policy justification for *res judicata* can be divided as follows:

a) Economic reasons – The courts resources are limited. The introduction of *res judicata* as an estoppel for relitigation permits the court to save its scarce resources, avoiding the use of public means in re-deciding disputes between the same parties. Through *res judicata*, the court can enhance its economic efficiency and provide for the rapid settlement of disputes.[23] This approach, although prior to the CPR introduction, is completely consistent with the overriding objective enshrined in CPR 1, i.e. the court must deal with the cases justly at a proportionate cost. This kind of justification aims at what Sorabji calls 'collective proportionality', balancing the court resources among all the users of the system.[24]

b) Stability of the decision – Additionally, *res judicata* plays a key role in the stability of and reliance in the adjudication system. The relitigation of disputes already settled by judicial authorities could lead to conflicting decisions on the same issues, thereby undermining the confidence of the general public in the civil justice system and decreasing its reliability.[25]

Avoidance of conflicting judgments has been treated as a similar feature present in *res judicata* and *stare decisis.*[26] Although *res judicata* does not possess general effects, both doctrines operate with the aim of stability and avoidance of the 'constant reconsideration' of that which has been already decided.[27] In fact, in common law, jurisdictions work together in the construction and development of the law, providing stability and confidence to the society through the upholding of their decisions.

[19] *Ibid.*

[20] Adrian Zuckerman (n8) 1239.

[21] *Ferrer v. Arden (1599)* 77 Eng Rep. 263, 266; 6 Co. Rep. 7 a (Eng.).

[22] Richard Shell, 'Res Judicata and Collateral Estoppel Effects of Commercial Arbitration ' [1988] 35(4) UCLA Law Review 640.

[23] *Ibid.*

[24] John Sorabji, *English Civil Justice after the Woolf and Jackson Reforms: A Critical Analysis* (Cambridge University Press 2014) 167–170.

[25] Adrian Zuckerman (n8) 1239.

[26] Robert von Moschzisker, 'Stare Decisis in Courts of Last Resort' [1924] 37(4) Harvard Law Review 409.

[27] *Ibid.*

Additionally, the courts are entitled to solve definitively the disputes presented to them, and those decisions are deemed to be final: if they are to be re-visited on request of one of the parties, the finality ends. In so doing, the system created for ensuring citizens the proper and fair vindication of their rights crumbles, and with it, the entire rule of law.[28] Through *res judicata*, and more precisely through its effects, the prestige and moral authority of judicial decision is protected[29]. Thus, *res judicata* becomes a gear of the civil justice system that allows people to respect the court's judgments as final, enhancing the stability of the entire system by informing them that the decision taken cannot be modified.[30]

Social stability and peace – Closely related to the preceding point is another important justification for *res judicata*, social peace.[31] Stability in the judicial decisions provided by finality in litigation opens the scope of *res judicata* to further social consequences.[32] Judgments issued by the court hold the unmatched characteristic of preclusion.[33] Through this preclusion, society understands that when adjudication has operated, the subject is inevitably settled.[34] Hence, *res judicata* enhances not only the stability of the system and confidence in the courts, but at the same time promotes certainty – certainty for the litigants in regard to their rights and in the claims presented for judgment.[35]

It is argued that this preclusive effect in *res judicata* prefers certainty to justice.[36] Once the court has given its opinion on the matter, the judgment becomes immutable, unless fraud or collusion is alleged,[37] even if new facts become available and new evidence is placed at the parties' disposal,[38] Thus, the finality of a mistaken judgment prevails over the justice sought even if this new evidence could be presented and the case re-litigated.[39] It is proposed that society and its citizens require the maximum level of certainty in their day-to-day relationships, and the confidence in the finality of jurisdictional resolution of disputes is a cornerstone of social stability.[40] In that regard, the stabilisation of rights vindicated in court through the immutability of the judgments helps the general public to understand that the disputes solved in court are, precisely, solved for good.[41] Hence, the conflict is seen as overcome, for no other litigation can be presented on the same grounds, and thus peace has been brought to the conflict, and consequently to society as well.

[28] Vaughan Lowe, 'Res Judicata and the Rule of Law in International Arbitration' [1996] 8(1) African Journal of International and Comparative Law 47–48.

[29] Richard Shell (n22). 641.

[30] Yuval Sinai (n7) 362.

[31] Zollie Steakley and Weldon Howell, 'Ruminations of Res Judicata' [1974] 28(1) SLJ 358.

[32] *Ibid.*

[33] Stephen Tromans, ' History that cannot be Re-Written: Res Judicata in Public Law' [1990] 49(2) The Cambridge Law Journal 192–194.

[34] Various Authors, 'Developments in Law Res Judicata' [1952] 65(5) Harvard Law Review 825–829.

[35] Zollie Steakley and Weldon Howell. (n31) 359.

[36] Laura Dooley, 'The Cult of Finality' [1996] 31(1) VULR 51.

[37] As pointed out by Lord Keith is Arnold "*In such a case the bar is absolute in relation to all points decided unless fraud or collusion is alleged, such as to justify setting aside the earlier judgment. The discovery of new factual matter which could not have been found out by reasonable diligence for use in the earlier proceedings does not, according to the law of England, permit* the latter to be re-opened". For a more detailed explanation on the concepts of fraud and collusion see Royal Bank of Scotland Plc v Highland Financial Partners LP [2013] EWCA Civ 328. Kuwait Airways Corporation v Iraqi Airways Corporation ("Perjury II") [2005] EWHC 2524 (Comm). Par. 197–199.

[38] *Ibid.*

[39] K. R Handley (n3) 6.

[40] Timothy Pinos, 'Res Judicata Redux' [1988] 26(4) Osgoode Hall Law Journal 718–719.

[41] Allan Vestal, 'Res Judicata/Preclusion by Judgment: The Law Applied in Federal Courts' [1968] 66(8) Michigan Law Review 1728.

PART II – ARBITRATION AWARDS AND DECISIONS UNDER THE AA1996

Arbitration as an alternative

The classic paradigm of civil justice has shifted over the last fifty years. Since the second half of the past century, alternatives to the traditional civil court adjudication model have flourished, particularly in common-law jurisdictions.[42] As part of this phenomenon, one of the most successfully understood and utilised alternatives has been arbitration, a court-like, private and simpler system which provides a solution without the costs, delays and vexations commonly linked to litigation.[43]

For the purpose of its development, Davidson's succinct yet clarifying definition of arbitration will be followed, according which arbitration is 'a process whereby two or more parties agree to submit a legal dispute to one or more third parties, whose role it is to pronounce judicially[44] on that dispute in the shape of a binding award'.[45]

From the case law, a recent definition by the court was given by Thomas J in Walkinshaw v Diniz, citing Hirst L.J. in O'Callaghan v. Coral Racing Limited, and it was stated that:

> To my mind the hallmark of the arbitration process is that it is a procedure to determine the legal rights and obligations of the parties judicially, with binding effect, which is enforceable in law, thus reflecting in private proceedings the role of a civil court of law. [46]

Thus, arbitration's distinctive features can be summarised by stating that the arbitrator is a third party who is privately chosen to 'make a decision within a procedural environment' who is also chosen by the parties.[47] The arbitrator's decision is to be considered final,[48] binding to the parties and enforceable in court.[49]

The quasi-judicial function of the arbitrator

Arbitration is, without doubt, the alternative dispute resolution mechanism that most resembles civil court litigation. Arbitration leads to a binding solution decided after an adversarial procedure of which the rules are,[50] in most cases, decided by the arbitrator according to the s. 34 AA1996.

The similarity to the judicial function undertaken by the arbitrator is recognisable in several rules in the Act that are not common to other ADR methods. Sections 100–103 allow and promote the enforcement of the arbitral award on similar grounds to a court's judgement and Subsection 69 allows, although exceptionally, the challenging of an award through an appeal in questions of law. According to this statutory recognition, arbitration leads to several of the judicial effects of judgements.

[42] Simon Roberts and Michael Palmers, Dispute Processes, ADR and the Primary Forms of Decision- Making (Cambridge University Press 2005) 1–2.

[43] David St John Sutton. (n2) 1.

[44] Judicially not as a synonym of court-like, rather as a manner to say that it is conducted through a proceeding. See Arenson v Casson Beckman Rutley & Co [1975] 3 All E.R. 901 at [914–915].

[45] Fraser Davidson, Arbitration (Westlaw UK 2015) 2.01. www.westlaw.co.uk. Viewed 8 August 2016 . Similar conceptualizations can be found in Shell, Richard Shell. (n22) 648, and Sime, Stuart Sime, 'Res Judicata and ADR' [2015] 34(1) CJQ 44.

[46] Walkinshaw v Diniz [2000] 2 All E.R. (Comm) 237, and the doctrine has been replied in England and Wales Cricket Board Ltd v Kaneria [2013] EWHC 1074 (Comm).

[47] Simon Roberts and Michael Palmers (n41) 264.

[48] David St John Sutton (n2) 1.

[49] *Ibid.*

[50] Sime (n44) 44.

Crucial differences between civil litigation and arbitration

Notwithstanding the aforesaid similarity, some important differences remain between formal civil litigation and arbitration. The relevance of noticing this dissimilarity resides in how they may be interpreted as an obstacle for further arguments regarding arbitration and litigation analogies. The most remarkable disparities are:

a) Appeals – Civil litigation encompasses a well-constructed statutory appellate procedure regulated in the CPR 52 and the Access to Justice Act 1999.[51] The scheme includes the possibility of first and second appeals to higher courts, and a filter device is allocated in the rule to avoid unnecessary appeals, namely permission to appeal.[52] Appeals perform an important double function for the civil justice system, correcting errors in particular cases and allowing the court to clarify and develop the law in appeals arising from question of law, furthering public confidence in the administration of justice.[53] For its part, in arbitration, as a general rule, appeals are very narrow under s. 67 to 69 of the AA1996,[54] based on 'serious irregularity', a lack of substantive jurisdiction or, exceptionally, as this is very restrictive and limited to English law, on questions of law.[55]

b) Precedent – Unlike civil justice, there is no precedent system applicable to arbitration, thus arbitrators are not bound by any prior decision made either by themselves or by other arbitrators regarding the same facts or the same issues.[56] The consequences of this is more flexibility for the arbitrator but also the possibility of conflicting decisions.[57]

c) Procedural and evidence rules – Litigation has an entire civil procedure code to deal with procedural matters. Furthermore, some other rules about specific topics are found in either other legal instruments – e.g. the Civil Evidence Act 1972 – or relevant court authorities, e.g. standard of proof rules. As part of its inherent flexibility, arbitration, does not follow strict rules of proceedings or evidence. The entire regulation of the matter is contained in s. 34 of the AA1996,[58] which grants the arbitrator broad powers to set the relevant regulation on this regard.[59]

d) Public function – The function of civil courts is to perform a constitutional obligation for the vindication of the rights of the litigants.[60] This duty is undertaken through the rules of a fair trial enshrined in Article 6 of the ECHR, applicable in the United Kingdom after the Human Rights Act 1998. In doing so, the court cultivates the confidence of the people subject to English jurisdiction in their power to definitively solve the civil disputes presented to them.[61] It has an essential role in the organisation of the state and the maintenance of the rule of law.[62] Arbitration, as a private dispute resolution mechanism, has no major concern either for public interest or for public confidence, and the arbitrator is solely a third person appointed for the resolution of a private and particular dispute.

[51] Adrian Zuckerman (n8) 1112.

[52] *Ibid.*

[53] *Ibid.*1113.

[54] David St John Sutton (n2) 531.

[55] Schwebel v Schwebel [2010] EWHC 3280 (TCC) (Tab 22) Akenhead J at [14]. Moreover, the rule under which the arbitrator decision can be appealed is waivable, as it can be agreed by the parties to opt out of it in their contract.

[56] Gabrielle Kaufmann-kohler, 'Arbitral Precedent: Dream, Necessity or Excuse? The 2006 Freshfields Lecture' [2007] 23(3) Arbitration International 357.

[57] Fraser Davidson (n44) at[2.45].

[58] (1) It shall be for the tribunal to decide all procedural and evidential matters, subject to the right of the parties to agree any matter. AA1996, s.34 (1).

[59] David St John Sutton (n2) 160.

[60] John Sorabji. (n24) 143.

[61] Hazel Genn, 'Understanding Civil Justice' [1997] 50(1) Current Legal Problems 160–165.

[62] *Ibid.*

e) Judicial authority – As a consequence of the aforementioned public function, the courts have the authority to compel individuals and public and private institutions in the compliance of the court's decisions, that is to say, they have coercive power.[63] The court, as a public authority, is entitled to restrain civil rights if required for the fulfilment of its mission. Courts are constitutionally authorised and even commanded to grant orders that limit the freedom of certain individuals or their assets, e.g. freezing orders, and failing to comply with court orders could lead to imprisonment (contempt of court). Arbitrators, as private individuals, do not possess such powers. For the fulfilment of their duties and the security of the arbitral proceeding, they require the assistance of the court,[64] for instance, with injunctions.[65]

PART III – THE APPLICABILITY OF *RES JUDICATA* TO ARBITRATION

The arbitral award is the final step of an adjudicative process, and its effects are binding for the parties regarding the conflict that has been resolved. However, the finality of that decision is not derived from the judicial authority, and thus the question about the applicability of *res judicata* to the award arises as one of maximum importance to the effectiveness of it, particularly in future (re)litigation.

As Sime explains, arbitration has been considered a sort of private version of court litigation,[66] thus the application of *res judicata*, at least to some extent, is not generally questioned. Hence, the aforementioned statement proposes two different questions, firstly whether *res judicata* applies to arbitral awards, and in the positive case, which reasons are provided for such an application.

The Applicability of Res judicata *to arbitral awards*

In English case law, as well as among scholars, the conclusion is apparently clear: *res judicata* seems to undoubtedly apply to awards decisions on very similar grounds to that which applies to judicial decisions issued by courts.[67]

This applicability tends to be treated as an assumption, yet the source of the application of the *res judicata* effect on arbitrational decisions is not as clear as the commonly given answer. Thus, Russell has suggested that both *res judicata* and abuse of process are applicable to arbitral tribunals' decisions,[68] specifying that 'a valid award will create an estoppel with regard to the matters with which it deals, preventing either party from pursuing those matters in a later stage of the arbitration or in subsequent proceedings',[69] an approach that has also been held on similar grounds in *Phipson*.[70]

The court, for its part, has established this application to arbitral decisions; in *Injazat Technology Capital Ltd v Dr Hamid Najafi,* Flaux J recognised the applicability of *res judicata* to an arbitral award toward avoiding re-litigation in other proceedings ruling: 'Accordingly, in my judgment, those are all claims either precluded by res judicata or

[63] *Ibid*. 155.

[64] s. 42 AA1996.

[65] s. 44 AA1996.

[66] Sime, (n44) 44.

[67] *Ibid*. See also Denise Bensaude, 'The International Law Association's Recommendations on Res Judicata and Lis Pendes in International Commercial Arbitration' [2007] 24(4) Journal of International Arbitration.

[68] David St John Sutton (n2) 358.

[69] *Ibid*.

[70] Hodge M. Malek, Phipson on Evidence (18 edn, Sweet and Maxwell 2013) 1497.

claims which are an abuse of process',[71] and a similar opinion was held in *Michael Wilson v Thomas Sinclair* and *Arts & Antiques Ltd v Richards.*[72]

Additionally, in *Lincoln National Life Insurance Co v Sun Life Assurance Co of Canada,* the Court of Appeal, provided a comparable suggestion by saying in the scope of arbitral proceedings:

> The principles of res judicata and issue estoppel apply between parties to the original proceedings or their privies.[73]

This aforementioned conclusion, it has been said, does not only apply to English arbitrations, rather, it extends to arbitral awards issued by arbitrators where the seat is located in foreign jurisdictions and enforced in England and Wales.[74]

Justification of res judicata *applicability in arbitral proceedings*

Although the conclusion appears to be clear, the reasons to confer the applicability of *res judicata* to arbitration have not been given either by the authors or in the referred case law by the courts. However, from the arguments provided by the literature, this work systematises the ideas in three doctrines. First, there are those who assign the *res judicata* effect on arbitral awards based on the similarity of the adjudicative process involved in arbitration and civil litigation; secondly some scholars have said that emanates from the statutory provisions of the AA1996. Conversely, there are others who insist that the source of the *res judicata* effect of arbitral decisions is not related to its likeliness to litigation at court, but rather its contractual nature, bounding the parties through their agreement and the submission to arbitration.

1) Similarity doctrine – The legal culture comprehends arbitration as the ADR method that resembles court adjudication the most, thus, the applicability of *res judicata* to it has been assimilated without this being considered exceptional or strange.[75] As long as the arbitral award has met the statutory requirements, the requisites contained in the agreement and the procedural rules agreed by the parties and the arbitrator, the *res judicata* effect of the decision is automatically assumed. Following this trend, the literature on this matter focuses its attention on the major preconditions of a valid arbitration, e.g. jurisdiction,[76] yet builds upon the belief that arbitration and litigation are similar enough to simply extend the application of *res judicata.*[77] As Shell critically commented, 'the courts seem to feel that arbitration is sufficiently like litigation to apply the same rules of preclusion to both types of proceeding'.[78]

The automatic transition from adjudication to arbitration of *res judicata* has been underpinned in several court decisions where the principle has been mentioned as applicable based on this implied similarity but with no major elaboration of the rationale underlying it. This supports Lord Denning, Master of Rolls' decision in *Fidelitas* Diplock L.J. to grant a *res judicata* effect to arbitral awards, who succinctly stated that: 'Issue estoppel applies to arbitration as it does to litigation.'[79]

[71] Injazat Technology Capital Ltd v Dr Hamid Najafi [2012] EWHC 4171 at [9].

[72] Michael Wilson v Thomas Sinclair [2012] EWHC 2560 at [49]–[50].

[73] Lincoln National Life Insurance Co Ltd v Sun Life Assurance Co of Canada [2004].

[74] Hodge M. Malek (n70) 1497.

[75] Sime. (44) 44–45.

[76] K. R Handley (n3) 54.

[77] Shell(n22) 657–658.

[78] *Ibid.* 658.

[79] Fidelitas Shipping Co. Ltd v. V/O Exportchleb [1966] 1 Q.B. 630. At 632.

More recently, in *Emirates Trading Agency v Sociedade de Fomento Industrial* Popplewell J. stated:

> The finality of the award creates an issue estoppel between the parties which precludes either party challenging it before the tribunal or as a ground of challenge to a subsequent decision of the tribunal.[80]

Furthermore, some authors have understood from the mentioned authorities that the court recognises the analogy among arbitration and litigation and have concluded that in *res judicata* matters, the 'same requirements apply to arbitral awards as to judgments',[81] reflecting the discussed conception of similarity or even identity of both mechanisms in the resolution of private disputes.

It appears from the aforementioned stance that the extension of *res judicata* to arbitration is not to be argued at all, being treated as sort of a natural consequence that flows from the evident similarity existing between arbitration and adjudication.

Critiques of the similarity doctrine

The aforementioned idea of an extension of *res judicata* to arbitration based almost exclusively on its similarity to adjudication could be strongly criticised, as expressed hereunder. The similarities between both institutions, although undeniable, are not as exact as the court and the authors appear to comprehend. As mentioned above,[82] arbitration and litigation have crucial differences, from their origin and purposes beyond the resolution of the particular disputes to the nature of the proceedings and the regulation of the decision-making process involved in each of them. As Gordon clearly pointed out, 'one cannot simply import preclusion from litigation to arbitration'.[83] The consequences of this identity rationale, or the lack of it, could lead to misconceptions and misguided conclusions toward the real effects of arbitration, all of them founded in what has been denominated 'the imperfect analogy' between arbitration and litigation.[84]

The imprecise application of the *res judicata* effect to arbitration, based on apparent similarities, is not only an argumentative error, but that very argument creates a risk of the misapplication of other institutions and privatives of public litigation to arbitration. This poses a threat to the administration of justice and the correct defence of the rights of the parties in arbitration and other ADR mechanisms. This peril is derived from the argumentative extension of the premise to further analogies: if the two methods in question, namely, litigation and arbitration, are so similar that the extension of *res judicata* is from one to the other is considered an obvious step, then it could eventually be concluded that both of them share more features and that more principles can be thus extended.

This conclusion contradicts the essence of both mechanisms of dispute resolution, infringing the distinctive and celebrated characteristic of arbitration – flexibility – as compared to adjudication. Based on the aforementioned differences between the two institutions, the application of this rationale could lead to a confused interpretation of some critical aspects of litigation in arbitral proceedings, all of them underpinned by the absence of a particular regulation in the rules of arbitration that makes applicable the litigation rules by an extended analogy:

[80] Emirates Trading Agency LLC v Sociedade de Fomento Industrial Private Ltd [2015] EWHC 1452 (Comm)at [23].

[81] Filip De lay and Audley Sheppard, 'ILA Interim Report on Res Judicata and Arbitration' [2009] 25(1) Arbitration International. 46.

[82] Section 2.3.

[83] Randy Gordon, 'Only One Kick at the Cat: A Contextual Rubric for Evaluating Res Judicata and Collateral Estoppel in International Commerce Arbitration' [2006] 18(2) Florida Journal of International Law 550.

[84] Shell (n22) 657.

a) Evidence and proceedings rules – The general rule stated in s.34 of the AA 1996 allows the arbitrator to control the proceedings. For the undertaking of his or her purpose, the statute grants broad powers to the arbitrator, who is entitled to set the rules regarding stages of the proceeding, times, terms and, very importantly, evidence rules. These powers are limited only by the limitations enshrined in the same rule regarding a prior or concomitant agreement of the parties on the rules[85]. Under the scope of an extended analogy between litigation and arbitration, a party could assume to be entitled to require the application of stricter procedural – e.g. CPR rules of the matter – particularly in subjects such as burden of proof or the cross-examination of witnesses. This behaviour could lead to abuse of the analogy by a party, toward trumping contractual obligations related to procedural rules contained in the arbitral agreement. Similarly, the invocation of the CPR rules could be used to force the arbitrator to withdraw or not apply the rules he or she established under s.34 at the beginning of the proceeding.

b) Appeals and Due Process – Subject to the same argument, the parties throughout the arbitral proceeding could attempt to expand the application of Section 69. They could attempt to do so by expanding the appellate system currently applicable in civil litigation under CPR and the Access to Justice Act to arbitral awards[86]. In fact, they even could misuse the similarity argument to present an appeal based on a breach of Article 6 of the ECHR, specifically, when the award is not accompanied by a major justification of the decision, stating that there has not been a decision on the merits and thus, there has not been a fair trial. Moreover, procedural fairness could be claimed to delay proceedings based, for instance, on a breach of the principle of due notice under rigid formal civil justice standards, inapplicable in a dispute resolution mechanism, which furthers confidentiality as one of its key features.[87]

The similarity rationale remains dubious, and in its unjustified application lies the danger of a completely chaotic and unwarranted assimilation of several principles of arbitration into adjudication. The rationale is oddly founded on the similarity between adjudication and arbitration, notwithstanding the fact that the relevance of arbitration as an alternative to adjudication lies in the differences between the two.

2) Statutory doctrine – Another interesting justification for the applicability of *res judicata* to arbitration is stipulated in English legislation, specifically in Section 58 (1) of the AA1996, which deals with the scope of the effect of an arbitral award.[88] The statutory disposition establishes that an award is 'final and binding', although no major definition of the meaning of the wording is given.

The reach of the expression 'final and binding' has major importance, and from the words chosen by Parliament, a strong foundation for the extension of preclusive effects of *res judicata* to arbitral decisions can be constructed. Notwithstanding the strength, at least *prima facie*, of the text included in the Act, is hard to find any important reference to it in most of *res judicata* literature regarding arbitration, and, when it is mentioned,

[85] It has been said that the arbitrator is the master of his own procedure, even though the parties can control such wide scope of freedom through their agreement (either on particular topics or to the entire arbitration) as stated both in subsection 34 of AA1996, which declares that: "It shall be for the tribunal to decide all procedural and evidential matters, subject to the right of the parties to agree any matter"; and letter f) of the rule, which adds: "whether to apply strict rules of evidence as to the admissibility, relevance or weight of any material". Notwithstanding that consensus possibility, in a majority of the cases the agreement would be towards a more flexible than a restrained or -court-like- approach. David St John Sutton (n2) 242.

[86] The parties tend to extend the s.69 through the concept of "question of law" Hew Dundas, 'Appeals on questions of law: section 69 revitalised' [2003) 69(1) Arbitration.9.

[87] Neil Andrews, Andrews on Civil Processes (Intersentia 2013) Vol. 2. 117.

[88] s. 58 Effect of award.
(1) Unless otherwise agreed by the parties, an award made by the tribunal pursuant to an arbitration agreement is final and binding both on the parties and on any persons claiming through or under them.

the focus is on the binding effect on the parties rather than the finality. The authors tend to centre the obligatory effect of the award and the entitlement of the parties to require the immediate enforcement of the decision.[89] The 'binding' effect is often the most commented on, interpreted as the enshrining of the mandatory force of the obligations created by the decision, replacing the parties' previous rights with those contained in the award,[90] and the effect of it on the arbitrator, who is bound by the award and his or her decision enclosed in it.[91]

Regardless of the aforementioned, some scholars have actually remarked on the relevance of Section 58, adducing that the word 'final' must be read as a definition of finality, stating that the award, 'is conclusive as to the issues with which it deals',[92] although no actual reference to *res judicata* is provided.

Fortunately for its part, the court has made a deeper development of Section 58 of the AA1996 and delved into the consequences of its introduction, dealing with it more closely connected to preclusion and finality. In *Hammad v Boltlake*, Eccles J, clarifying the expression used in s. 58, stated:

> By virtue of s.58 the arbitrator's award was binding and could not be set aside. (2) There was clearly an issue estoppel that arose out of the arbitration proceedings. Arnold v National Westminster Bank Plc (No.1) [1991] 2 A.C. 93 applied.[93]

A similar opinion was given in *Sheffield United FC v West Ham United FC,* where Teare J said:

> An award of the arbitral tribunal is final and binding on the parties in the sense that the decision of the arbitral tribunal alone will finally and exclusively determine the issues between the parties.[94]

More recently in Gloster J made an effort to clarify the meaning of the expression by stating that:

> I agree that the word "conclusive" does not alter the position, and that a phrase such as "final, conclusive and binding" in the context of an arbitration agreement such as clause 14 of FIA does no more than restate what has long been the rule in relation to arbitrations, namely that an award is final, conclusive and binding in the traditional sense, in that it creates a res judicata and issue estoppel.[95]

All of these judgments provide useful guidance due to the fact that in both, although no express mention of *res judicata* is given, issue estoppel, which is a preclusive effect of the *res judicata* doctrine, is assumed. Thus, the court is interpreting that the wording of the AA1996, namely 'final and binding', engenders preclusive effects to awards.

In the international arbitration field, there is a comparable absence of justification beyond the discussed similarity. Foreign arbitral awards are valid in England and Wales through their recognition and enforcement as enshrined in Sections 100–103 of the AA1996 and should encompass the same preclusive effects of domestic awards. In this context, it has been held that the international instruments that have produced

[89] David St John Sutton (n2) 351.

[90] *Ibid.*

[91] Gbangola v Smith and Sheriff Ltd [1998] 3 All E.R. 730 Lloyd J at [738].

[92] David St John Sutton. (n2) 351.

[93] Hammad v Boltlake Ltd [2010] QBD Unreported.

[94] Sheffield United FC Limited v West Ham United FC Plc [2008] 2 C.L.C. 741(QBD (Comm)Teare J at [14].

[95] Shell Egypt West Manzala GmbH v Dana Gas Egypt Ltd [2009] EWHC 2097 (Comm) Gloster J at [38]. The main issue of the case relied on whether the words "final, binding, and conclusive" were meant no exclude the possibility of an appeal under s. 69.

model laws for arbitration would have recognised the application of the *res judicata* to awards.[96] Thus, the UNCITRAL model law, on which the AA1996 is, to some extent, based,[97]establishes in Articles 32(2) and 35(1) that the award must be 'final and binding'.[98] This mention of finality and mandatory effects of the award to the parties has been interpreted as a formal recognition to *res judicata* in the literature,[99] yet it has been the courts which have provided further clarification in the leading case of *Shell Egypt v Dana Gas Egypt Ltd,* in which the rule of UNCITRAL Article 32(2) was included by the parties in the arbitral agreement as governing law. Gloster J stated:

> The reality is that the expression "final and binding", in the context of arbitration, and arbitration agreements, has long been used to state the well-recognised rule in relation to arbitration, namely that an award is final and binding in the traditional sense and creates a res judicata between the parties. [100]

As seen in the cited case law, both domestically and internationally, there is a recognition of the finality and preclusive effects of the award as a consequence of the interpretation given to the expression 'final' of the indicated statutes.

Critiques of the statutory doctrine

Notwithstanding the fact that the Section 58 of the AA1996 entails a better justification than simply the similarity between arbitration and litigation, the explication on its own is not sufficient. The expression 'final' seems to have a broad significance and it is hard to determine the real scope of it. Additionally, the interpretation suggested by the court, although helpful and resolute, fails to provide a reason for the application of *res judicata* to arbitration beyond stating that it actually applies, basing their conclusions on 'traditional sense' and in the 'well recognised rule' that attributes estoppel effects to the words 'final and binding'.

Furthermore, there is another critical matter that is not treated in the cited authority: the nature of the rule contained in s.58 of the AA1996. That is the rule of a 'non-mandatory provision' according to s.4(2) of the AA1996.[101] That is to say, as Russell suggests, the parties can agree to alter the effects of the award contained in s.58, including the 'final and binding' consequences.[102]The right for the parties to opt out of non-mandatory rules is based on the very private nature of arbitration; hence, parties can always agree to prevent a particular rule from applying to their dispute, as far as the rule at stake is not mandatory. Furthermore, *res judicata* is built as an estoppel for the party who is facing relitigation, and as such, it is always up to them to plea it and prove it, rather than them being entitled to having the court to act ex oficio on their behalf[103]. That means that *res judicata* is always disposable for the parties holding the power to decide whether to trigger it or not as part of their defences at court. From this perspective, the rule contained in s.58 should not be seen as a major issue, however it is noteworthy that it relies on what the statute acknowledges is non-critical role for

[96] Filip De lay and Audly Sheppard (n84) 60.

[97] Andrew Tweeddale and Karen Tweeddale, A Practical Approach to Arbitration Law (Blackstone 1999) 35.

[98] Andrew Tweeddale and Karen Tweeddale, A Practical Approach to Arbitration Law (Blackstone 1999) 35.

[99] Filip De lay and Audly Sheppard (n84) 60.

[100] Shell Egypt West Manzala GmbH v Dana Gas Egypt Ltd [2009] EWHC 2097 (Comm) Gloster J at [38].

[101] (1) The mandatory provisions of this Part are listed in Schedule 1 and have effect notwithstanding any agreement to the contrary.
(2) The other provisions of this Part (the 'non-mandatory provisions') allow the parties to make their own arrangements by agreement but provide rules which apply in the absence of such agreement.

[102] David St John Sutton. (n2) 351.

[103] See N° 3 infra.

finality, a flexible approach to *res judicata*, assuming that 'final and binding' refers to its application, from a legislative perspective. For the lawmaker, the estoppel derived from an award is an option; something for the parties to ultimately decide, rather than an unfailing consequence of the resolution of the dispute by a third party with a binding effect for the parties. In other words, the preclusive consequences of the award are different than their litigation counterpart, and therefore, the standpoint toward *res judicata* is to be different; it cannot simply be constructed from the wording of the Act, nor from the simple analogy of arbitration to adjudication. The distinction is not trivial, because what is mandatory and unavoidable in civil litigation becomes a mere possibility in arbitration; thus, the entire rationale of *res judicata* appears not to be fully applicable[104].

3) Contractual doctrine – Among scholars there is a doctrine that aims to explain the preclusive effects of the arbitral award from a *ius privatum* perspective. As mentioned above, it is habitual for the literature to centre on the similarity between arbitration and litigation as the main criteria for the expansion of *res judicata* from the latter to the former. Often not mentioned in this analogy is the critical difference in the collective interest behind *res judicata*, a public interest that is not present in arbitration. As Gloster J commented in *Shell*, finality in arbitration 'springs from a different source';[105] being a creature of contract, the correct justification, he suggests, must be sought in arbitration's contractual nature rather than in public interest considerations. In arbitration there is no general concern about the rational use of court resources, nor any preoccupation regarding the overriding objective of the civil justice system; the only relevant resources to be taken into account are those belonging to the parties.

The contractual doctrine does not deny the necessity of conclusiveness; this doctrine acknowledges a *res judicata* equivalent in arbitration but addresses the foundations from a distinct perspective. When a dispute arises, each party holds the right to bring a claim to court,[106] hence, by the inclusion of an arbitral agreement, are excluding that right by a contractual term as part of the negotiation of the main contract. This means that parties renounce their constitutional right to attend court in exchange for the other rights and obligations comprehended in the bargaining of the contract,[107] and the preclusive effect of the eventual arbitral award is one the relevant issues negotiated. All of the arbitration owes its existence to the arbitral agreement,[108] thus, "the parties' expectations regarding preclusion are thus an integral part of the arbitration agreement and should be the primary focus of any preclusion analysis'.[109]

Therefore, the arbitral agreement as part of the contract is binding for the parties, but for the arbitrator and the courts as well. In Posner's words, finality is 'attached to arbitrator as a matter of contract rather than as a matter of law, the preclusive effect of the award is as much creature of the arbitration contract as any other aspect of the legal-dispute machinery established by such a contract'.[110] Thus, the rationale for preclusion associated to the arbitration process resides in the binding force and enforceability of contracts, at least in everything that is covered under the scope of the agreement.[111]

[104] See N° 3 Infra.

[105] Shell (n22) 662.

[106] Shell (n22) 662.

[107] *Ibid.*

[108] Randy Gordon (n82) 559.

[109] *Ibid.*

[110] Richard Posner giving judgment in Ids Life Insurance Company and American Express Financial Advisors, Inc., Plaintiffs-appellants, v. Royal Alliance Associates, Inc., et al., Defendants-appellees, 266 F.3d 645 (7th Cir. 2001).

[111] Randy Gordon (n82) 559.

Consequently, the contractual doctrine states that the parties control the preclusive effects of the award and are able to adjust the scope of the finality. Parties can agree in advance to the arbitral clause that the award would have a limited estoppel effect, restricted to only some of the treated issues, establish a limitation of time for the estoppel to be operative for future litigation, and more drastically, to exclude the finality from the award. All of these possibilities would be encompassed in English Law under the non-mandatory provision of s.58(1), which contains the provision 'unless otherwise agreed by the parties', permitting the modification of the natural effects of the award and their adaptation to a tailored form of preclusion.

Criticism of the contractual model

Although the contractual model provides a novel explanation for the existence of the preclusive effect of arbitral awards, separating it from the traditional rationale of *res judicata* in adjudication, several criticisms of it can be made.

Firstly, it is plainly apparent that the doctrine is problematic on the grounds that it assumes that arbitration agreements are always freely negotiated. The aforementioned creates an explicative vacuum toward statutory arbitration, whereby parties are not free to choose whether or not to introduce an arbitral agreement or to select the effects of that arbitral proceeding or the preclusive consequences of the award[112]. Statutory arbitration is not completely uncommon under English and Welsh law, and it is present in several regulated industries, as happened in agriculture with the Agricultural Holdings Act 1986,[113] the construction of roads with the New Roads and Street Works Act 1991,[114] and industry under the Industry Act 1975.[115] In all of these cases, the fundamental premise of the contractual model for preclusive effects to arbitration could not apply, for it was founded in party autonomy, a private law principle that was trumped by statutory imposition.

Secondly, from a pure contractual perspective, the model fails to properly explain the effects of arbitral awards when the parties are relatively silent in the arbitral agreement. As Hulbert asserts, most of the time, parties do not include an explicit clause to deal with the preclusive effects of the potential arbitration.[116] The parties, he claims, never seek to define in their contracts the clear preclusive and conclusive effects of arbitration.[117] Thus, courts have to face a major problem to interpret what, supposedly, the parties' intention may be according to the "contextual evidence", putting an "almost unbearable" burden upon courts.[118] Furthermore, this interpretative work would cause delay and vexation, leading to a disproportionate usage of court resources in a sort of 'satellite litigation', particularly if is taken into account that the main issues of the dispute would have already been settled in arbitration.

[112] Moreover, the contractual-based doctrine tends to accept the doubtful premise of party equality and symmetrical position to entering the negotiation process.

[113] Schedule 6 of the Act.

[114] Pt III, that refers to the application of the AA1996 on the matter.

[115] s.20 AA1996.

[116] Richard W. Hulbert, 'Arbitral Procedure and the Preclusive Effect of Awards in International Commercial Arbitration ' [1989] 7(2) International Tax & Business Lawyer at 198.

[117] *Ibid*. 199. A different approach can be found in Born, who states that the mention to preclusive effect is not necessary "[i]t is inherent in the nature of an agreement to arbitrate, and the concept of an arbitral award, that such an award will have binding, and thus preclusive, effects". Gary Born, *International Commercial Arbitration* (2nd edn, Wolters Kluwer 2014) 3745.

[118] *Ibid*.

Effects of the award – issue estoppel?

As noted in Section 1.2,[119] the effect of *res judicata* in English law, as cited from *Arnold,*[120] is composed by cause of action estoppel and issue estoppel. The importance and application of the former have been expressed above; hence, this section addresses the potential inclusion (or not) of issue estoppel in arbitration, given its characteristic features of being more subtle and harder to apply, even in civil litigation.[121]

Issue estoppel is a subsidiary rule of *res judicata,*[122] i.e. due to the structure of *res judicata*, it is only applicable when claim estoppel cannot be configured to be applicable to the case. It impedes the re-litigation of issues that have or ought to have been settled by a previous decision where that particular issue is to be considered as fundamental and central to the dispute.[123]

Because arbitration is a contractual mechanism to resolve private disputes, the application of issue estoppel demands a more specific analysis. This is especially true, considering that disputes submitted to arbitration are often well defined and, in not a few cases, thoroughly circumscribed in the arbitration clause of the contract or the arbitral agreement.[124] Thus, it is relevant to disclose the interaction of all the issues that arise throughout the arbitration proceedings, how they are contained in the arbitral award and whether they exceed, or not, the context of the arbitral agreement. Hence, it could be determined how the inclusion of those issues in the final decision could lead to an issue estoppel to a further litigation, considering the limited scope of the arbitration process when compared to civil adjudication.

Issue estoppel and the award

General approach

The question of whether issue estoppel is applicable to arbitrator's awards seems to have a relatively clear answer as mentioned above, however the relevance of the consequences of the discussion make it compelling to delve into the pertinent literature a little more deeply.

Following the general trend regarding the application of *res judicata* to arbitration,[125] the English scholars and judicial authorities have acknowledged the applicability of issue estoppel to arbitral awards. This recognition has often been hidden under the broad umbrella of *res judicata* and its extension to arbitral proceedings, under which it must comprehend both the cause of action estoppel and of issue estoppel. Nonetheless, there have been noteworthy mentions regarding issue estoppel specifically, in an effort to provide a clearer and clarifying interpretation. Thus, *Russell* suggests that issue estoppel is to apply according to the general principles on the same grounds that it applies in court judgements,[126] an opinion shared in *Redfern and Hunter,*[127] as well as in other jurisdictions.[128]

[119] Section 1.1.

[120] See note 7.

[121] Allan Vestal (n6) 860.

[122] Ghosh Narun, 'An uncertain shield: res judicata in arbitration' [2015] 31(4) Arbitration International 663.

[123] *Ibid.*

[124] Nigel Blackaby, Redfern and Hunter on International Arbitration (Thomson Reuters) at [2.29–2.32] www.westlaw.co.uk. Viewed 8 August 2016.

[125] Section 3.1.b.

[126] David St John Sutton. (n2) 356–358.

[127] Nigel Blackaby, Redfern and Hunter on International Arbitration (6th edn, Oxford University Press 2015) 561.

[128] As it happens in the US, where issues are fully and fairly adjudicated in a prior arbitration. Similarly, the ILA endorsed the inclusion of issue estoppel in its recommendations included in the ILA Final Report on Res Judicata in Arbitration. Redfern and Hunter (n120) 561.

From the judicial perspective, the leading authority was given by Diplock LJ in the cited case of *Fidelitas,*[129] and the same doctrine was held by the Privy Council in *Associated Electric v European Reinsurance Co of Zurich*, where the court said that issue estoppel was: 'a species of the enforcement of the rights given by the [previous] award.'[130]

This rationale was more recently endorsed in the leading case of *Arts & Antiques Ltd v Richards,*[131] where Hamblem J gave a clarifying opinion on a specific issue by saying:

> I am quite satisfied that an issue estoppel arises in respect of the Arbitrator's decision that the insurance contract was subject to a policy wording which contained CP2. [132]

As shown above, the courts and scholars have agreed that it appears that the English law of *res judicata* applies to litigation as it applies to arbitrations.[133]

Dissenting opinions

Being consistent with their idea of a justification for *res judicata* in arbitration diverted from the analogy with formal adjudication,[134] contractual doctrine advocates suggest that issue estoppel should be handled carefully.[135] As arbitration is a private arrangement, the court or arbitral tribunal entitled to apply issue estoppel shall focus on the contractual intention of the parties. Thus, the court must determine, according to the principle of contractual construction of the arbitration agreement, whether the parties had or had not borne in mind the particular preclusive effect alleged, meaning, whether they wanted or not to be bound, not only by the main decision, but by the fact-finding of the arbitrator regarding the issues treated before him.[136] Consequently, if the application of issue estoppel for a particular issue encompassed in the arbitration is not consistent with the contractual concerns of the parties, then the estoppel should not be applied.[137]

Some scholars have criticised the applicability of issue estoppel to arbitration, to the extent of suggesting that its scope should be extremely reduced and restricted only to specific cases where the parties expressly stipulated the estoppel.[138] Thus, unlike Shell,[139] the task of the second tribunal or court is not to reconstruct an implied agreement of the parties, rather, it must only apply when it is undoubtedly clear from the arbitral agreement.[140]

The foundation for this argument relies on the consequences that issue estoppel brings to arbitration, where it is considered somehow potentially dangerous.[141] The broad application of issue estoppel to arbitration would place an additional burden upon arbitrators: the burden of making a decision with judicial-like effects without the

[129] Note 60, supporting the plain application of res judicata to arbitration.

[130] Associated Electric & Gas Insurance Services Ltd v European Reinsurance Co of Zurich [2003] 1. W.L.R. 1041. Lord Hobhouse at [15].

[131] Mark Beeley and Hakeem Seriki, 'Res judicata: recent developments in arbitration' [2005] 8(4) International Arbitration Law Review 112.

[132] Arts & Antiques Ltd v Richards [2013] EWHC 3361 at [25].

[133] Mark Beeley and Hakeem Seriki, (n128) 116.

[134] Section 3.1.3.b.

[135] Shell (n22) 667.

[136] *Ibid*. 668–669.

[137] *Ibid*.

[138] Hiroshi Motomura, 'Arbitration and Collateral Estoppel: Using Preclusion to Shape Procedural Choices' [1988] 63(1) Tulane Law Review 79.

[139] Notes 109–110.

[140] Hiroshi Motomura. (n131). 80.

[141] Increasing the level of justification of arbitrator's fact –finding. *Ibid*.

judicial formality of proceedings and decision-making.[142] Consequently, this would force arbitrators to decide the cases with judicial standards. The consequences – issue estoppel – of their conclusions would affect other future cases through the preclusion of certain issues, even if the case is not closely connected to the award, allocating unnecessary pressure to the arbitral machinery.[143]

Accordingly, the application of issue estoppel would be forcing, again, the analogy between arbitration and formal adjudication.[144] Thus, it does not acknowledge that court adjudication fulfils a social function that goes much further than that of solving civil disputes; it executes other functions, what Motomura calls supplemental law-making effects,[145] which, in aggregate terms, are even more important than the particular dispute, as collateral – issue estoppel being one of them. Conversely, he suggests that these supplemental functions are not, nor should they be, present in arbitration.[146] Arbitrators would be obliged to raise the formality, paradoxically affecting arbitration by increasing the similarity with adjudication and its drawbacks.

Hence, issue estoppel is suggested only as a complementary doctrine. Unlike cause of action estoppel, issue estoppel 'is neither an essential nor even a general attribute of dispute resolution' and, if applied as a general effect of arbitral awards, would undoubtedly affect the arbitrator's freedom.[147] Thus, this trend leads to a reduction in the relative attractiveness of arbitration, risking its position as a real alternative to litigation.[148]

Relevance of issue estoppel application to arbitration

So far, this essay has looked at the reasons provided to support the application of issue estoppel in arbitration as well as the contrary opinion and the main arguments. It is important, however, to establish the consequences of such a discussion beyond the argumentative debate. Two concrete consequences of the application of issue estoppel are particularly relevant for the analysis and are strongly related with the nature of arbitration. These are, first, the caution that must be taken with non-arbitrable issues, and second, the problem that arises when issues are arbitrable but fall outside the scope of the arbitral agreement.

Non-arbitrable issues

Although the AA1996 does not expressly deal with non-arbitrability, there is general consensus on the fact that certain topics cannot be solved in arbitration; generally, these topics are related with state matters, public policy,[149] and specific subject matters excluded from arbitration by statute, where a provision states that the parties hold an inalienable right to access to the courts.[150]

There are some subject matters where issues are not clearly determinable, particularly in claims where some issues are arbitrable and others are not. In these cases, the preclusive

[142] Hiroshi Motomura. (n131). 79.

[143] *Ibid.*

[144] *Ibid.*

[145] This refers to the function of civil courts in the justice system, encompassing rights vindication and law development through precedent.

[146] *Ibid.* 80–82.

[147] *Ibid.*

[148] *Ibid.* 81.

[149] O'Callaghan v Coral Racing Ltd (CA) [1998] All ER (D) 607.
The arbitrability was denied as the arbitration agreement was part of a betting contract that was declared void under the Gaming Act Legislation.

[150] David St John Sutton (n2) 71.

effects of the awards must be examined carefully. For a more accurate analysis, a division can be established regarding the source and clarity of the non-arbitrability. First, if the issues are not arbitrable because of a clear statutory provision, then the award should not be granted with any issue estoppel effect nor any preclusive consequence to the issue that should have never been included in the arbitral proceedings. Similarly, any conclusion provided by the arbitrator on the issue should be devoid of any further effect. [151]

On the other hand, if the same issue encompasses both arbitrable and non-arbitrable parts of the claim, or if the provision in the statute is not clear enough about the topics concerned in the arbitration, the courts must decide which effects are going to be granted to the awards regarding those issues. In doing so, the courts should be cautious, particularly when arbitrability – on its own – is a discussed topic.[152] An expansive approach to the issue estoppel effect of these awards could impact the arbitrability decisions of the court. That is, if courts decide to give properly preclusive effect to awards that contain non-arbitrable issues, it could create a trend toward a more restrictive approach to arbitrability; that is, a tendency to decide in favour of non-arbitrability in dubious cases as a method of protecting the integrity of the system.[153] Unfortunately, the consequence of that could lead, at least theoretically, to, , an atmosphere of distrust toward arbitration.

Issues excluded from the arbitration

Another topic closely connected with non-arbitrability is the scope of the arbitration agreements. The mere interpretation of the agreement toward the determination of the actual scope of the arbitration usually entails a relatively high number of court judgments[154] and all the cost and delay associated with civil litigation for the sole purpose of establishing the arbitral framework.[155]

The concern lies in those issues that are not included, or are expressly excluded, in the arbitral agreement and that, for different reasons, are actually disputed in the arbitral proceedings.[156] In this case, the issues, or the whole claim, is statutorily arbitrable, yet the parties decide to restrict the scope of the arbitration to only certain disputes arising, currently, or in the future, from the contract.[157] The exclusions agreed by the parties commonly have the purpose of diverting some remedies from formal adjudication,[158] e.g. the restriction in the arbitration clause to only contractual damages, leaving all other damages and remedies excluded.[159]

[151] If the issues were the central facts of the claim, then there is not only a lack of issue estoppel effect, there is no res judicata effect whatsoever because of the absence of jurisdiction by the arbitrator.

[152] David St John Sutton (n2) 71.

[153] Hiroshi Motomura, 'The Relationship Between Arbitration and Court Litigation in the United States' [1989] 39(4) Hokkaido University Collection of Scholarly and Academic Papers 401–398.

[154] See e.g., Investors Compensation Scheme v West Bromwich Building Society [1998] 1 All E.R. 98.
Paul Smith v H&S International Holding Inc [1991] 2 Lloyd's Rep. 127.

[155] David St John Sutton (n22) 71. That is to say, initiate proceedings in court for the sole reason of fixing the actual matter falling under the arbitration. As Russell points out "is the type of dispute that has arisen within the scope of the particular arbitration agreement? This will be the focus of the following paragraphs. Thirdly, even if disputes of the type in question are within the terms of the arbitration agreement, does the dispute fall within the scope of what was actually referred by the parties in the particular reference?"

[156] *Ibid.*

[157] This division of claims have led to important litigation in England, as seen in Guidance Investments Ltd v Guidance Hotel Investments Co BSC [2013] EWHC 3413 (Comm) where it was disputed whether a gross negligence counterclaim was part of the arbitration agreement or not.

[158] What Schwartz denominates as remedy-stripping. David Schwartz, 'Understanding Remedy-Stripping Arbitration Clauses: Validity, Arbitrability, and Preclusion Principles' [2003] 38(1) University of San Francisco Law Review 50–55.

[159] *Ibid.* 90–94.

Hence, although the scope of the arbitration may be narrow and most of the possible issues must still be litigated at court, the findings of the arbitrator could still lead to further preclusive effects in subsequent court proceedings.[160] The predicament arises when the parties introduce to the former arbitral proceedings claims that comprehend 'hybrid' issues,[161] or issues that do form part of the arbitral agreement and upon which the arbitrator at least partially makes the decision. If one of the parties seeks to rely on these findings by an issue estoppel in a subsequent proceeding, should the court acknowledge that estoppel?

The question could be addressed from two different perspectives, and in both, the simpler, but not definitive, answer would be negative. First, from a more procedural approach, if the parties excluded certain claims or limited the scope of the arbitration, then the arbitrator has no jurisdiction on the topics beyond that scope. Accordingly, the results of an award that exceeded the arbitrator jurisdiction should not lead to issue estoppel, and the party affected by the results should be entitled to properly challenge the award under Section 67(1) of the AA1996,[162] notwithstanding the use of Sections 32 and 73 in their respective proper timing.[163]

Secondly, from a contractual focus, considering arbitration as a creature of contract, the award should not engender any preclusive effects on the topics that fall outside the scope of the arbitration. As Shell suggests, the parties excluded from arbitration a claim that they were willing to dispute before the court, thus, their intention was precisely to not give any *res judicata* effects to them, except from the effect derived from court adjudication.[164]

Parties, therefore, should be obliged to not include into arbitration issues not incorporated in the arbitration agreements as a matter of contract. However, it seems necessary to remark that, if contractual doctrine applies here, then the contractual principles at stake could allow a reinterpretation of the arbitral agreement according to the conduct of the parties throughout the arbitral proceedings. An example of the aforementioned would arise when a party is seeking to avoid the issue estoppel in subsequent proceedings, but the same party implicitly accepted the introduction of formerly excluded issues in the prior arbitration, and even benefited from it, e.g. by the enforcement of the award that contained the problematic issues. In the opinion of the author of the current work, the court could interpret, according to the contractual doctrine, that the real intention of the parties toward the arbitration was implicitly modified. Thus, the award would lead to preclusive effects and the issue estoppel would be sustained.

PART IV – PROPOSED APPROACH, RECOMMENDATIONS AND CONCLUSIONS

After the explanation and proper criticism of the doctrines that aimed to provide a justification for the applicability of *res judicata* to arbitration and their further consequences, this essay will undertake the statement of the author's own position. This position encompasses a broad-based justification in support of the application of *res judicata* to arbitration in which all of the above explanations contribute to some extent.

[160] *Ibid.* 94.

[161] David St John Sutton (n2) 38.

[162] Peter Aeberli, 'Jurisdictional Disputes under the Arbitration Act 1996: A Procedural Route Map' [2005] 21(3) Arbitration International 265–275.

[163] Sections 32, 32, 67, and 73 are all mechanisms for the parties to object and challenge the tribunal's jurisdiction.

[164] Shell (n22) 670. Also see Shell Egypt West Manzala GmbH v Dana Gas Egypt Ltd [2009] EWHC 2097 (Comm) Gloster J at [38].

Primarily, the immediate justification for the extension of *res judicata* to arbitral awards on its two branches – cause of action and issue estoppels – emerges normatively from s.58 of the AA1996. As mentioned above,[165] the wording 'final and binding' has been uniformly interpreted by the court as synonymous of finality, and the pertinent authority has understood that the effects aimed by the provision are those of *res judicata*. However, as mentioned in the corresponding section, the exclusive–normative explanation is not enough.[166]

It is additionally necessary to understand the role that arbitration plays nowadays in the broad dispute resolution framework in England, and, more specifically, in its close relationship with civil litigation. Although both litigation and arbitration have distinguishing features that impede an automatic analogy between them, it must be borne in mind that arbitration is, in actuality, the ADR mechanism that most requires court adjudication, and, consequently, has the greatest impact upon it. In fact, the AA1996 is constructed to locate the court, although with restrained interference,[167] in a supervisory position. Thereby, this allows the arbitrators to request the court's assistance – particularly its coercive powers – to permit them to carry out their tasks, e.g. freezing injunctions[168] and summoning witnesses[169] throughout the entire arbitral proceeding.[170]. Furthermore, the court, in accordance with s.72 of the AA1996, is entitled to determine the jurisdiction of the arbitrator, an issue which often leads to a significant number of litigations in English courts. Moreover, as mentioned in Part III, the arbitrability itself of the subject manner is often decided by the court, and this has a direct impact on the very existence of the arbitral proceeding.[171] Similarly, the compulsive enforcement of the award under s. 66 of the AA1996 is restricted to the constitutional powers exclusively granted to courts, and even the finality discussion, such as the proper *res judicata* effect of them in subsequent proceedings, is, as Sime explains, a highly litigated topic.[172]

The aforementioned examples illustrate the complex and cooperative relationship between arbitration and court adjudication – a 'marriage', in Andrews' words[173] – and the complex way in which arbitration's satellite disputes may have an effect on the workload of the courts. The point is not about their similarity; conversely, the AA1996 has considered the courts as complementary to the arbitration.[174]

Thus, arbitration engenders the use of the civil courts' resources, and that is far from insignificant. It is a strong demand of scarce means which, oddly, are destined to be what is said to be the 'alternative' to formal adjudication. As a consequence of this, the preclusive effects of arbitration find their rationale in the public concerns of *res judicata*. Thereby, *res judicata* in arbitration has also a dual function, a private one related to the finality of the decision, allowing the parties to rest and move on, but it additionally fulfils a function of public interest, on very similar grounds to civil litigation. The public rationale of arbitration is that it is intended to avoid the relitigation of arbitral issues in

[165] Notes 79–90.

[166] See "Critiques to the statutory doctrine" in Part III.

[167] As stated in Part 1 of the AA1996, the main principle is non-intervention, unless a provision requests the court assistance or the exercise of its jurisdiction. David St John Sutton. (n22) 7–8.

[168] s. 44 of the AA1996.

[169] s. 43.

[170] Okezie Chukwumerije, 'Judicial Supervision of Commercial Arbitration: The English Arbitration Act of 1996' [1999] 15(2) Arbitration International 191.

[171] Note 139.

[172] Sime (n44) 36.

[173] Neil Andrews (n88) 112.

[174] Okezie Chukwumerije (n166) 190.

court and to further social stability. The aforesaid point is particularly relevant when a finalised arbitration includes several visits to court, e.g. over jurisdiction and arbitrability matters. In these cases, the *res judicata* works as a defence of litigation and the civil justice system as a whole from the disproportionate utilisation of public resources in a dispute of private resolution. This means that if the award is not granted with *res judicata*, at least with cause of action estoppel, the risk of a second proceeding – either arbitral or in court – would constitute a repetition and a waste of expenses, time, and result in the vexation of the parties and the delay of the system.

This approach entails a justification for *res judicata* in arbitration that is consistent with the overriding objective enshrined in CPR 1, allowing arbitration to use only a proportionate part of the court means. Thus*, res judicata* stands as a closure rule that concurrently supports both arbitration and litigation. It supports arbitration by providing a preclusive system that makes it reliable and conclusive, and litigation by preventing the resurrection of the 'unholy trinity of civil justice', which is based, as Andrews suggests, [175] on the endemic problems of every justice system in the world: cost, time and stress; hence, underpinning the avoidance of duplicative, costly and vexatious proceedings and preventing inconsistent decisions.[176]

However, for this justification to work optimally the interpretation of s.58 must be more restrictive, impeding the contractual modification of the preclusive effects of the award. If the parties were entitled to withdraw the finality from the arbitration, the public purpose of *res judicata* would be in danger. Thus, the finality of arbitration should not be considered as a completely non-mandatory provision.

CONCLUSION

This work has explicated the relevance of *res judicata* and its applicability in arbitration as well as the main arguments provided in by judges and scholars for a proper justification of that applicability. In that task, this essay has presented three different doctrines, the similarity doctrine, the statutory doctrine and the contractual doctrine. All of those three justifications are here considered insufficient. Thus, the article has delved into the necessity of a wide perspective which includes civil litigation at court and arbitration as complementary and not only as alternatives, due to the constant relationship existing between both mechanisms, particularly on issues such arbitrability, scope of the agreement, injunctions and enforcement. Then, a broader approach is proposed, stating that the correct justification of *res judicata's* application to arbitration lies in private, but also public concerns, related with the overriding objective of civil justice and the public purpose of *res judicata* derived from the cooperative relation that exists between arbitration and litigation under the regulation of the AA1996. Additionally, it has been stated that the scope of the *res judicata* doctrine applicable to arbitration should encompasses, invariably, cause of action estoppel, albeit regarding issue estoppel caution must be taken with non-arbitrable issues, and issues that fall outside the scope of the arbitral agreement.

[175] Neil Andrews, 'A new civil procedural code for England: party-control 'going, going, gone'' [2000] 19(1) CJQ 20.

[176] Silja Schaffstein, The Doctrine of Res Judicata Before International Commercial Arbitral Tribunals (Oxford University Press 2016) 117.

BALANCING FUNDAMENTAL RIGHTS IN PRIVATE LAW THROUGH THE DOUBLE PROPORTIONALITY TEST

*JOHN T CHEUNG**

ABSTRACT

Rights review in public law typically pits the rights of the individual against the interests of the government. However, the recent introduction of fundamental rights into private law raises the question as to how competing rights-based claims may be reconciled, given that both parties are able to plead the protection of fundamental rights in support of their respective positions. This article proposes that the 'double proportionality' test, which enables the interference with the rights of both parties to be assessed respectively to determine whether it is proportionate, represents the most promising solution to this difficulty. The test allows the courts to give sufficient attention to the rights of all involved, and just as importantly, enables courts to be seen as doing so. The article will also examine the jurisprudence on this area and criticise the courts' failure to apply the double proportionality test satisfactorily (or at all) in conflicting rights cases.

Keywords: double proportionality; balancing fundamental rights; constitutionalisation of private law; horizontal effect of fundamental rights.

INTRODUCTION

We live in an age of balancing. In the context of rights review in public law cases, balancing typically pits the rights of the individual against the interests of the government. It is a familiar process in which only one party – that is, the private claimant, as opposed to the public defendant – is generally able to invoke the protection of fundamental rights. However, the introduction of fundamental rights into private law in recent years – sometimes referred to as the 'constitutionalisation' of private law[1] – raises the difficult question as to how competing rights-based claims may be reconciled, as both parties are potentially able to plead the protection of fundamental rights in support of their respective positions.

In a nutshell, this article contends that the 'double proportionality' test, which enables the interference with the rights of both parties to be assessed respectively to determine whether it is necessary and appropriate, represents the most promising solution to this difficulty. The article is divided into two sections. First, I will argue that, in principle, the test is commendable for it allows the courts to give sufficient attention to the rights of all involved, and just as importantly, enables courts to be *seen* as doing so. In other words, its use is necessary to achieve substantive and procedural justice for both parties. Secondly, the focus of the article will shift to an analysis of the cases (primarily English decisions), where the double proportionality test has hitherto been confined to the context of cases involving a clash between privacy and the freedom of expression. Regrettably, it will be seen that the courts have time and again failed to apply the double

*BA (Hons 1st Class) (Oxon), BCL (Dist) (Oxon). Pupil barrister at Temple Chambers, Hong Kong. I am grateful to the Hon Mr Justice Andrew Cheung, Chief Justice of the Court of Final Appeal of Hong Kong. All errors are, of course, my own.

[1] M Kumm, 'Who's Afraid of the Total Constitution? Constitutional Rights as Principles and the Constitutionalization of Private Law' (2006) 714 German Law Journal 341.

proportionality test satisfactorily in such cases. Worse still, there is no mention of the double proportionality test *at all* in cases falling outwith this narrow domain.

Before proceeding, I should clarify that the article focuses solely on conflicting rights in the context of private law (particularly tort, contract and property), as opposed to public law. By private law, I am referring generally to the body of law which governs relationships between private actors (individuals, groups, organisations, and corporations), as opposed to relationships between citizens and the state. The argument will assume the indirect horizontal applicability of human rights law in private law claims.[2] However, the distinction between public and private law is a thin one. [3] Some of the 'private law' cases discussed below (such as those reaching the Strasbourg court) will involve public defendants. This is due to the modern recognition that public authorities are not only under a negative obligation to respect fundamental rights, but also a positive obligation to protect rights against infringements by third persons, including private actors. For present purposes, such claims will be categorised as private law cases insofar as the substance of the disputes stems from direct violations of fundamental rights by private actors.

The argument is also methodological rather than theoretical. It examines the appropriate analytical tool which courts should employ when balancing competing rights in private law disputes. Accordingly, it falls beyond the ambit of my thesis to consider the vexed issue as to the foundation or general theory of fundamental rights; for the sake of clarity, this article will proceed on a specific understanding of fundamental rights, namely an understanding which perceives fundamental rights as 'special' norms that hold presumptive priority over public interests.[4]

THE ARGUMENT IN PRINCIPLE

The merits of the double proportionality test

Fundamental rights have traditionally been perceived as restricted to the domain of public law. Yet, it is now increasingly apparent that they can, at the very least, shape the development of private law, if not be directly applicable against other private parties.[5] The acceptance of fundamental rights discourse in private law claims raises a particular difficulty. As a private individual, one is entitled to a sphere of individual sovereignty, which has been carved out to the benefit of each individual by the government or legislature. But problems arise when one individual's sphere of sovereignty intersects with another's sphere of sovereignty: it then becomes necessary for the court, as guardians of our fundamental rights, to resolve the conflict. Sometimes, the dilemma is avoided by 'defusing conflicts as fake',[6] *viz* by finding that a right is not engaged, or in drawing

[2] There is a wealth of legal literature on whether and, if so, how Convention rights were to affect relationships between private parties governed by common law. See, eg, D Oliver and J Fedtke (ed), *Human Rights and the Private Sphere: A Comparative Study* (Routledge-Cavendish 2007); D Friedmann and D Barak-Erez (ed), *Human Rights in Private Law* (Hart Publishing 2001); D Hoffman (ed), *The Impact of the UK Human Rights Act on Private Law* (Cambridge University Press 2011). As regards the positive law, it would appear that English courts have not accepted the proposition that Convention rights enjoy 'direct horizontal effect', namely the view that individuals can directly rely on Convention rights to make claims against other private individuals (see most recently *Akhter v Khan* [2020] EWCA Civ 122). It is nevertheless clear that the courts have been willing to develop the existing common law in the light of Convention rights, thus tacitly favouring an 'indirect horizontal effect' approach (see, eg, *Campbell v MGN Ltd* [2004] UKHL 22, [2004] 2 AC 457).

[3] W Lucy, *Philosophy of Private Law* (OUP 2007) 12–25; D Priel, 'The Political Origins of English Private Law' (2013) 40(4) Journal of Law and Society 481.

[4] This specific understanding of fundamental rights is contested, however. See S Smet, *Resolving Conflicts between Human Rights: The Judge's Dilemma* (2017) 29–31.

[5] Eg *Campbell v MGN Ltd* (n 2).

[6] E Brems, '*Evans v UK*: Three Grounds for Ruling Differently' in S Smet and E Brems (eds), *When Human Rights Clash at the European Court of Human Rights: Conflict or Harmony?* (OUP 2017).

the boundaries of individual rights in such a way as to avoid the conflict.[7] At other times, however, it will be necessary to evaluate the strength of each claim to individual sovereignty when a clash occurs, in other words, to undertake the task of balancing competing rights. However, the rhetoric of 'balancing' conceals more than it reveals. In a broad sense, balancing can be loosely defined as an economic, or cost-benefit, analysis of law. In the context of rights review, however, both English and Continental European courts have favoured a more structured balancing exercise, namely the proportionality test.[8] The question then becomes whether this test – more precisely labelled as the *single* proportionality test – is applicable in the private law context to resolve clashes between fundamental rights.

There is a crucial way in which this task differs from how courts have dealt with fundamental rights claims against public authorities, as powerfully articulated by Collins.[9] In public law cases, the question is typically whether the government's case for the need to override a right in the pursuit of a compelling public interest is established. In a private law dispute, however, the balancing exercise assumes a rather different flavour by dint of the fact that both parties can claim that their fundamental rights are at stake, and the issue is how to measure those rights against each other. An example of the latter would be a case involving a clash between the fundamental right of a property owner (A) to retain control of his property, and the fundamental right of a protester (B) to express herself through making use of the property as a site of protest. Those rights imply a correlative legal duty: B would have to refrain from certain action if A's right is deemed of greater weight and the interference with B's right is regarded as proportionate; or, alternatively, A would have a duty to tolerate (or no power to restrain) B's right, if that was adjudged to be of greater weight, and the infringement of A's right proportionate.

Collins posits that this structure prevents the application of the familiar public law test of proportionality, because this transplant will not function to provide a procedure by which all the different relevant considerations are measured against each other. This leads him to conclude that a different test, namely the 'double proportionality' test, ought to be adopted in such cases, since it would allow the interference with the rights of both parties to be assessed respectively to determine whether it is necessary and appropriate. Lord Steyn fleshed out the details of the test as follows in the seminal House of Lords case of *Re S*:[10]

> First, neither Article has *as such* precedence over the other. Secondly, where the values under the two Articles are in conflict, an intense focus on the comparative importance of the specific rights being claimed in the individual case is necessary. Thirdly, the justifications for interfering with or restricting each right must be taken into account. Finally, the proportionality test must be applied to each.

Pausing here, more ought to be said about the double proportionality test. Whilst Collins has made a persuasive case for abandoning the classic single proportionality test in cases concerning conflicting rights, this still leaves unanswered the question of why the double proportionality test should be utilised instead. This is particularly

[7] S Smet (n 4) 19.

[8] The proportionality test consists of four stages: (i) whether its objective is sufficiently important to justify the limitation of a fundamental right; (ii) whether it is rationally connected to the objective; (iii) whether a less intrusive measure could have been used; and (iv) whether, having regard to these matters and to the severity of the consequences, a fair balance has been struck between the rights of the individual and the interests of the community.

[9] H Collins, 'On the (In)compatibility of Human Rights Discourse and Private Law' in H-W Micklitz (ed), Constitutionalization of European Private Law (2014) 26.

[10] *Re: S (Identity: Restrictions on Publication*) [2004] UKHL 47, [2005] 1 AC 593 [17].

important given that, as demonstrated in Section 3 below, the courts have not generally applied the double proportionality test satisfactorily (or at all) in such cases, and have at times treated the single proportionality approach as a panacea for all thorny cases requiring the balancing of rights and interests. This may well be prompted by a lack of awareness of the limitations of the single proportionality test, but could also be due to an ignorance as to the strengths of the double proportionality approach.

In my view, the most compelling justification is rooted in considerations of fairness and justice. The double proportionality test allows the courts to give sufficient attention to the rights of all involved, and just as importantly, enables courts to be *perceived* as doing so. In competing rights cases, the imperative of doing justice to both parties entails a separate consideration of each purported limitation of a fundamental right. Apart from elucidating to the losing party the basis upon which the court has favoured his or her opponent's right, the double proportionality test also improves the chances of sound conclusions being reached, by pushing the court to work through the respective cases for limiting either right. If one right emerges as the clear victor whichever analysis is used, the outcome would be straightforward. But if a different conclusion is reached by the court depending on which right was afforded initial priority as the primary right, the strength and clarity of the two analyses must be scrutinised.

All this deserves some elaboration. From a dignitarian viewpoint, the double proportionality test secures procedural justice in cases involving conflicting rights for it acknowledges not only the existence of competing rights involved, but also the fact that those rights are all deserving of the full protection of the proportionality test, which *ex hypothesi* seeks to minimise any proposed interference with the right in question. In contrast, applying the single proportionality test to a case involving conflicting rights has the propensity to prioritise one set of rights over the other – and even if the same result is likely to be achieved under the double proportionality test, the party who 'loses' under this approach (and whose rights have not been subject to a separate proportionality analysis) is likely to feel aggrieved due to the perception that his rights have been neglected. The value of the test is on this view dignitarian and non-outcome: there is an intrinsic benefit in applying a separate proportionality test to the rights of not only the claimant but also the defendant, because respect for the individual demands that he be treated fairly, and his rights preserved to the greatest extent possible. After all, justice must not only be done, but also seen to be done. A failure to use the double proportionality approach in such cases undercuts the sense of fair play on which our legal system is founded, and sends the message that some rights are better protected than others.

Is the issue simply one of perception of fairness? Is there any substantive merit to the double proportionality test? At first blush, it might be thought that the two proportionality tests should yield the same outcome, and that it matters not which proportionality test is applied. In other words, it is essentially the same question being examined from two perspectives, rather than two distinct tests. Put another way still, the task is to ascertain an equilibrium – to restrict one right to the extent necessary to protect a competing right – and the two tests should give the court the balance point. The double proportionality test is, on this view, purely symbolic and rhetorical in nature and function; it serves no instrumental purpose (*viz* it does not substantively influence the outcome of such cases). One might illustrate this view by way of a case concerning a 'kiss and tell' newspaper story, which typically involves an attempt to restrain the publication of private information through an injunction. Such a case would see a conflict between A's freedom of expression and B's privacy. Applying the double proportionality test, the court would first ask whether the protection of A's freedom of expression justifies an

intrusion on B's privacy, and whether that intrusion is necessary and proportionate; it would then assess whether the protection of B's privacy justifies an intrusion on A's free expression, and determine whether the intrusion is necessary and proportionate. If the privacy right demands the restriction on free expression, it is said, this must mean that the privacy right cannot be restricted because of free expression and hence no restriction would be proportionate. What renders the restriction on free expression proportionate is that it is necessary to protect the right to privacy. It makes no difference whether one adopts the single or double proportionality test.

When the basis of this line of reasoning is scrutinised, however, it is found to be wanting. It neglects the crucial point that the single proportionality test is liable to favour and give presumptive weight to the right being examined under the test (right A), by seeking to minimise any intrusion on that right at the third stage of the test, at the expense of the competing right engaged (right B). The very structure of the proportionality test, and the third stage in particular, implies that an interference with a right can only be considered lawful if it is no more than is necessary to achieve a particular legitimate aim. The right is therefore given utmost protection, as against any competing right or interest, as the courts must consider whether there is a less restrictive alternative. Indeed, it is precisely this feature of the proportionality enquiry – described by Tremblay as the 'priority-to-rights' principle[11] – which has resulted in its emergence as the key bulwark against the abuse of human rights by the government. As Smet elaborates, the 'priority-to-rights' model ensures that 'the invoked human right is granted principled (but defeasible) priority over the considerations invoked to justify its infringement'. Protection of the invoked human rights is the norm; justified restrictions are the exception. The single proportionality test is therefore loaded in favour of the invoked human right. This is demanded by the special normative force of fundamental rights and perfectly coherent when a fundamental right is opposed by a public interest in a public law claim. But Smet's thesis also evinces that, in a *private* law case concerning conflicts between human rights, the 'priority-to-rights' model 'unjustifiably skews' the court's analysis in favour of the directly invoked fundamental right (right A), and downplays the gravity of the competing fundamental right at stake (right B). The court is prompted to frame the case around the claimant's right A while the defendant's right B threatens to 'fade into the background'.[12]

Now, it is true that the fourth stage of the proportionality test – the balancing stage – compels courts to weigh rights A and B against each other, but it bears stressing that in order to reach this final stage, one must first go through the preceding three stages. The danger is that the fourth stage will not even be reached. In seeking to maximise protection of right A in the third stage, the court may well conclude that there is a less intrusive alternative which may be used to protect right B. There is an unacceptable asymmetry here, as right A is subject to this 'maximisation' or 'prioritisation' treatment, whereas right B is not. Asking whether right B can be protected *only* through a particular measure and not any other is a distinct question from whether right B can be protected *at all*. If, then, the single proportionality test is employed in a case where the relevant rights are pulling in different directions, the fear is that such an approach would fail to take both parties' rights with equal concern and respect.

The point may be tested through the 'kiss and tell' example above. If the court were to apply the single proportionality test only in relation to, say, A's freedom of expression,

[11] L Tremblay, 'An Egalitarian Defense of Proportionality-Based Balancing' (2014) 12 I•CON 864; cf M Klatt, 'An egalitarian defense of proportionality-based balancing: A reply to Luc B. Tremblay' (2014) 12(4) I•CON 891.

[12] S Smet (n 4) 36.

the first two stages would be easily satisfied: it is tolerably clear that the protection of privacy can, in theory, constitute a 'legitimate aim' which can justify an intrusion on the freedom of expression; similarly, the proposed injunction is plainly rationally connected to that objective. However, regarding the third stage, is there a less restrictive measure which the court can employ to protect B's privacy? Instead of granting an injunction, it is possible that privacy can be given *some* degree of protection through other less drastic measures, such as damages in lieu. It is not at all axiomatic that the granting of an injunction is the *one and only* measure which can serve to uphold B's privacy right. Self-evidently, damages in lieu would not accord nearly as much protection as an injunction – but notice that the court here is not required to maximise protection of privacy, but only protection of freedom of expression. It can be forcefully argued that applying a separate proportionality test to the privacy right is necessary to counteract this inherent bias and ensure that both rights are being treated as normative equals.

Admittedly, applying the single proportionality in a competing rights claim will not perforce result in the wrong outcome. The single proportionality test is elastic and need not be implemented mechanically. In the 'kiss and tell' example, it is conceivable that a court could simply conclude that, given the status of the right to privacy as a fundamental right, it *is* absolutely necessary to grant an injunction (and thereby to restrict A's freedom of expression) in order to substantiate the legitimate aim of protection of privacy. No lesser alternative (such as damages in lieu) will do. Applying the single proportionality test is not problematic, it might be reasoned, as judges can simply give due weight to the privacy right by framing the legitimate aim appropriately, and applying the third stage of the proportionality test such that the privacy right is also given maximum protection.[13] However, even granting the premises, such an approach is dissatisfying. It demands a distortion of the single proportionality test, for the third stage is not applied in a way which maximises protection of the right to freedom of expression (as it should), but instead in a way which implicitly maximises protection for *both* rights. If so, one might as well apply the double proportionality test for reasons of transparency and clarity. Although correct results may well be arrived at under the (modified or distorted) single proportionality test, the double proportionality approach provides judges with a superior and more lucid mode of resolving disputes between two right-holders. Ultimately, one should not forget that judges are only human. Consideration of the same issue from two separate perspectives is conducive to sound judicial decision-making, as judges are reminded of the need to reach a balanced view which fairly takes into consideration the value and importance of both rights involved. At the very least, it serves to improve the odds of a correct balance being struck.

Three further observations

Put these points to one side. Three further observations on the merits of the double proportionality test bear mentioning. In the first place, as a matter of logic, the application of the double proportionality test flows naturally from the acceptance of the single proportionality test as the correct approach when courts undertake rights review in public law challenges. Put differently, if we accept the premise that the single proportionality

[13] This is the approach adopted in Hong Kong in cases concerning clashes between the freedom of expression and private property rights: *HKSAR v Fong Kwok Shan* (2017) 20 HKCFAR 425 [69]. But note the recent developments in *Cheung Tak Wing v Director of Administration* [2020] 1 HKLRD 906, where the Hong Kong Court of Appeal acknowledged at [45] that a double proportionality approach might be adopted to resolve any conflict between free expression and property rights (however, on the facts, it was held that there was no issue of competing rights to resolve).

test is apposite where a claimant is challenging a governmental decision for infringing his or her fundamental rights, it must follow that, where rights are invoked by both litigants, the double proportionality test should be used to assess the proportionality of a purported interference with the respective rights.

The second observation is closely linked with the first one. In terms of legal obligation, English courts are now obligated to apply the proportionality test when assessing the legitimacy of an interference with both Convention and common law rights.[14] The same goes for derogations from European Union legal rights.[15] Even more pointedly, *Re S* is direct authority for the proposition that the double proportionality test should be employed in conflict of rights cases.[16] A failure to apply the double proportionality test in competing rights cases would therefore signal a striking abdication of judicial responsibility. The authorities confirm the necessity to ensure that all rights must be protected via the proportionality test, and this entails a separate application of the proportionality test to all the rights engaged in any given case.

Thirdly, employing the single proportionality test in a conflicting rights case would render the application of the test contingent upon the happenstance of the identity of the right-holder bringing the claim. In a public law challenge, the court is accustomed to applying the proportionality test vis-à-vis the *applicant's* rights, which is understandable given that the public defendant is (generally) not able to invoke any fundamental rights to defend the claim. Such a mindset would be ill-suited to private law claims raising conflicting rights. It would mean that, if a landowner were to bring a claim to evict protesters, the court would apply the single proportionality test in relation to his property rights; whereas if protesters were to seek a declaration that they could lawfully protest on the landowner's land instead, the same test would now be applied to assess the proportionality of the interference with their right to protest. There is a real danger that such an approach would give priority to the applicant's right for the simple reason that it has been invoked by the applicant, a phenomenon described by Smet as 'preferential framing'.[17] This renders the outcome of judgments dependent on an arbitrary factor, namely the identity of the right-holder initiating legal proceedings.

Applicability of the double proportionality test

The bulk of the article thus far has concentrated primarily on the third and fourth propositions of Lord Steyn's formulation of the double proportionality test in *Re S*, that is, the necessity for judges in conflicting rights cases to consider the justifications for interfering with or restricting both rights involved, and to apply the proportionality test separately. The impression given is that, whenever the court is faced with a conflicting rights claim, the double proportionality test must be resorted to. However, one crucial caveat should be noted. Recall that Lord Steyn's formulation of the double proportionality test specifically stressed that, as a precondition for the application of the test, neither article can have as such precedence over the other. The implication is that, if one article does enjoy priority over another competing article, no such balancing would be necessary. One can only speculate as to what his Lordship had in mind, but it is likely that his Lordship was referring to the distinction between absolute rights and non-absolute

[14] *Smith and Grady v UK* (1999) 29 EHRR 493; *R (Daly) v Secretary of State for the Home Department* [2001] UKHL 26, [2001] 2 AC 532.

[15] The Treaty on European Union (art 5(4)).

[16] *Re S* (n 10) [17].

[17] S Smet (n 4) 35.

rights under the European Convention of Human Rights.[18] The Convention reserves the label 'absolute' to a small number of rights, including the rights not to be tortured and not to be held in slavery or servitude. This supports the view that absolute rights are an *exception* to proportionality analysis, as opposed to the *result* of proportionality analysis.[19] That is to say, proportionality evaluates when the infringement of a fundamental right is justified, with such justification being available only if there is a limitation clause providing for it. If no such clause exists, then there can be no justification for infringing a right. This reading chimes with Strasbourg's understanding of absolute rights,[20] and has the weighty support of Möller[21] and Barak.[22] It is also buttressed by conflicting rights cases within the property context. In the South African case *Victoria and Alfred Waterfront v Police Commissioner of the Western Cape*,[23] the Western Cape High Court, faced with an application to ban the respondents permanently from begging on the commercial premises of the applicants, ruled that:

> The issue of begging frequently raises a direct tension between the right to life and property rights. In that event, the property rights must give way to some extent. The rights to life and dignity are the most important of all human rights. By committing ourselves to a society founded on the recognition of human rights we are required to value those rights above all others.

In determining the limits of the property right, the Court undertook a demarcation exercise and not a balancing of the relevant rights. For obvious political, social and above all moral reasons, a property right is not weighed against the right to life.[24] A similar example is the Supreme Court of New Jersey decision in *State of New Jersey v Shack*,[25] which, on the Court's view, saw a direct clash between the exercise of the landowner's property right (to exclude) and the migrant workers' right to life and human dignity. Again, the Court treated this as a matter of ascertaining where the boundaries of the property right have to be drawn to secure the life and dignity rights; no balancing was deemed necessary.

A SURVEY OF THE CASES

Having unpacked the reasons why the double proportionality approach should be adopted when balancing rights in private law cases, I now move on to survey the case law on this area, which can be neatly divided into two main categories: (i) those concerning a conflict between art 10 (freedom of expression) and art 8 (privacy), and (ii) those concerning a conflict between other rights.

[18] Another possible interpretation is that civil and political rights would have priority over economic and social rights. However, this is questionable since human rights instruments do not provide a hierarchy between fundamental rights. Other than the division of 'absolute rights' and 'relative rights', there is no indication of a ranking among fundamental rights. Indeed, the Canadian authorities have consistently guarded against a hierarchical approach to rights, see, eg, *Dagenais v Canadian broadcasting Corp* [1994] SCR 835 (SCC), [31]. But see now A Stuart, 'Back to basics: without distinction – a defining principle?' E Brems (ed), *Conflicts between Fundamental Rights* (Intersentia 2008) 101–130, who argues that the 'non-distinction norm' should have priority over another fundamental right with which it might compete.

[19] G Webber, 'Proportionality and Absolute Rights' in V Jackson and M Tushnet (eds), *Proportionality: New Frontiers, New Challenges* (CUP 2016).

[20] *Chahal v United Kingdom* Application No 22414/93, Merits and Just Satisfaction, 15 November 1996 [79]–[80].

[21] K Möller, *The Global Model of Constitutional Rights* (OUP 2012) 180.

[22] A Barak, *Proportionality: Constitutional Rights and their Limitations* (CUP 2012) 35.

[23] [2004] 1 All SA 579.

[24] AJ van der Walt, 'The Modest Systemic Status of Property Rights' (2014) 1 Journal of Law, Property, and Society 15, 51. Indeed, the right to beg arguably bears particular significance when one bears in mind the welfare system in South Africa, and how it contrasts with the welfare systems in some of the Organisation for Economic Co-operation and Development ('OECD') countries.

[25] NJ 297 (1971).

Conflict between art 10 and art 8

Cases which expressly apply the double proportionality test have largely been confined to the domain of tort law, particularly the competition between art 10 and art 8 in breach of confidence (or 'misuse of private information') claims[26]. This is perhaps not unexpected: even though *Re S* itself was not a breach of confidence claim, the case concerned a conflict between those very two rights, and in laying down the double proportionality test, Lord Steyn specifically cited the leading breach of confidence case *Campbell v MGN*.[27]

Do these cases demonstrate a proper application of the double proportionality test? Rather ironically, it would seem that Lord Steyn himself failed to apply his own test, at least in a fashion which accords equal weight to both rights, to the facts of *Re S*. The case arose from a court order by a family judge prohibiting the identification of the name or school of a 5-year old boy and preventing any publication in a report of the criminal trial of the name or photograph of his mother or his deceased brother. The gagging order was made for the welfare and protection of the boy whose brother was murdered allegedly by their mother. Several newspapers intervened to ask for a variation of the gagging order. The judge acceded to the application and relaxed the order, and the House of Lords dismissed the boy's appeal. Having articulated the details of the double proportionality test, Lord Steyn then endorsed the balance struck by the judge (in favour of art 10), observing that:

> Given the weight traditionally given to the importance of open reporting of criminal proceedings it was in my view appropriate for [the judge], in carrying out the balance required by the [Convention], *to begin by acknowledging the force of the argument under article 10 before considering whether the right of the child under article 8 was sufficient to outweigh it.* (my emphasis)[28]

The problem with this line of reasoning is that the attention seems to be solely on the importance of art 10, and whether privacy is sufficiently important to outweigh it. Is this really a decision that balances both rights according to a double test of proportionality? The judge's analysis, which Lord Steyn upheld, adopts as a starting point the 'primacy'[29] of art 10 due to the importance of open reporting of public proceedings on grave criminal charges, and then asks whether art 8 can place any limits on the full extent of the reporting. This looks very much like an application of art 10, particularly art 10(2) which provides for derogations from art 10(1). Indeed, the judge's conclusion specifically stressed that:

> Last, I am simply not convinced that, when everything is drawn together and weighed, it can be said that grounds under article 10(2) of the [Convention] have been made out in terms of the balance of the effective preservation of [the child]'s article 8 rights against the right to publish under article 10.[30]

If the court unduly favoured art 10 at the expense of art 8 in *Re S*, the converse is true in *McKennitt v Ash*.[31] The case concerned a well-known folk music recording artist, who claimed that a substantial part of the book written by her former friend revealed personal and private details about her that she was entitled to keep private. The judge

[26] J Morgan, 'Privacy, Confidence and Horizontal Effect: "Hello" Trouble' (2003) 62 CLJ 444.

[27] *Campbell* (n 2).

[28] *Re S* (n 10) [37].

[29] *Re S* [2003] EWHC 254 (Fam) (First Instance) [19].

[30] *Ibid*.

[31] *McKennitt v Ash* [2006] EWCA Civ 1714, [2008] QB 73.

granted an injunction preventing further publication of a significant part of the work complained of on the ground that it constituted private information under art 8, and held that her right under art 8 outweighed her former friend's right to freedom of expression under art 10. On appeal, it was submitted that the judge had failed to apply the proportionality test separately to each of the rights, but this was rejected by the Court of Appeal on two grounds.[32] First, the Court of Appeal espoused the view that the submission conflicted with the express reasoning of the judge, particularly the following passage:

> I need naturally to consider each of the passages in the book singled out for complaint separately, not only to decide whether in each case the threshold test for privacy is passed (that is to say, whether or not there would be a reasonable expectation of privacy), but also to consider, if that initial test has been satisfied, whether any other 'limiting factor' comes into play such as public domain or public interest.[33]

This is perplexing. If anything, the quoted reasoning of the judge corroborates, rather than contradicts, the appellant's submission, as the impression given is that once the right to privacy had been made out, the question would then become whether any considerations should serve to qualify it (eg public domain or public interest). Indeed, this is evident in the Court of Appeal's own discussion of the public interest in disclosure later on in its judgment. Referencing Strasbourg's decision in *Von Hannover v Germany*,[34] Buxton LJ found that, even if the artist could be said to be a 'public figure' in the relevant sense, there were no 'special circumstances' to justify or require the exposure of her private life.[35] It is striking that the whole analysis of his Lordship stays within art 8, without any proper balancing of the two rights.

The second reason relied on by the Court of Appeal was that, '[t]he suggestion that the judge, having so directed himself, needed none the less to repeat that direction as a mantra every time he came to a specific issue is quite unreal'.[36] But this neglects the point, made in the first section, that applying the proportionality test separately to the individual rights involved is rooted firmly in considerations of justice. Far from constituting needless repetition, the double proportionality test constitutes what is necessary to satisfy the need to afford the rights of both parties the appropriate attention they deserve.

Viewed thus, it is plain that there is a serious deficiency with the way the courts have gone about balancing rights in breach of confidence cases. Whilst paying lip service to the double proportionality test, the courts have in truth refrained from a proper consideration of both rights engaged when resolving conflicts between them. Nevertheless, it is possible that both *Re S* and *McKennitt* were rightly decided on the facts. Any interference with fundamental rights must be measured against the strength of the legitimate aims advanced. In *Re S*, Lord Steyn observed that the child's art 8 claim was relatively weak, given that the child would not be involved in the trial as a witness or otherwise; it would not be necessary to refer to him; no photograph of him would be published; and there would be no reference to his private life or upbringing. Any impact on his art 8 right caused by the reporting of his mother's trial would be, in his Lordship's words, 'essentially indirect'.[37] The restriction of art 8 was thus

[32] *Ibid* [48].

[33] *McKennitt v Ash* [2005] EWHC 3003 (QB) (First Instance) [67].

[34] (2005) 40 EHRR 1; (2006) 43 EHRR 7.

[35] *McKennitt* (n 31) [59].

[36] *McKennitt* (n 31) [48].

[37] *Re S* (n 10) [25].

proportionate to the objective of upholding the competing art 10 right, and there was accordingly no need to restrict art 10 through the gagging order in order to prevent the legitimate goal of preserving art 8 from being unacceptably compromised. Similarly, in *McKennitt v Ash*, the Court of Appeal paid fairly close attention to the question how far the book really did invoke the former friend's art 10 right to tell her own story (that includes her various experiences with the folk artist). While the former friend had been involved in some of the matters revealed, and a spectator of many others, the Court stressed that the book was not in any real sense about her at all: '[The former friend] gives vent to many complaints about the [celebrity]; but the interest of those is that they are complaints about the [celebrity], and not at all that the complaints are made by the [former friend]'.[38] Indeed, it is doubtful whether the former friend's art 10 right to tell her own story was engaged at all, given the Court of Appeal's view that the former friend had no story to tell that was her own as opposed to being the celebrity's.

Since the courts in both cases arguably got the answer right, it might seem churlish to complain of its reasoning. But the advantage of recognising that the courts were engaged in the task of balancing fundamental rights, and that the double proportionality test ought to have been applied in full, is that it directs attention to both sides of the claims. In purporting to apply the double proportionality test, the courts actually implemented the single proportionality approach, and it is submitted that this severely dulls the courts' analyses in both instances.

Conflict between other rights

A more fundamental worry concerns the courts' reasoning in cases falling outwith the breach of confidence context. After all, the potential for rights to compete with each other clearly exists in other areas of private law – contract, tort, and property – once one assumes the horizontal effect of human rights (a premise which is admittedly contentious). It is discouraging to observe a complete absence of any mention of the double proportionality test as a balancing mechanism to resolve conflicts between fundamental rights in this broader domain, and thus a consequential failure to accord equal respect to the rights of both parties, as I shall now explain.

In *Bull v Hall*, the Supreme Court held that the Christian hoteliers' refusal to grant double-bedded accommodation to a homosexual couple, on the basis that their religion regarded sexual relations between all unmarried couples (whether heterosexual or homosexual) as immoral, amounted to unlawful discrimination on the ground of sexual orientation.[39] For present purposes, what is pertinent is that this prompted a submission, advanced on behalf of the hoteliers, that this protection of the rights of the gay couple by antidiscrimination laws amounted itself to an illegitimate restriction of the right of the hoteliers' to manifest their religion. The Supreme Court was quick to reject this. The facts of the case provide a classic example of a case between private parties involving a clash of fundamental rights: the homosexual couple was entitled to be protected against discrimination on the basis of protected characteristics, for those rights ensure individual dignity and equal respect, as well as their right to respect for private and family life (right A), but equally the hoteliers were entitled to be able to freely choose a contractual partner, in addition to other important rights such as freedom of religion and the right to exclude unwelcome people from their private property

[38] *McKennitt* (n 31) [31].

[39] (2013) UKSC 73.

(right B).[40] In holding in favour of the homosexual couple, Lady Hale laid emphasis on the fact that the right to respect for their sexual orientation constitutes a core component of a person's identity under art 8.[41] Her Ladyship underlined the continuing legacy of centuries of discrimination and persecution against homosexuals which is still going on in many parts of the world, and concluded that 'very weighty reasons' would be required to justify discrimination on grounds of sexual orientation.[42]

Pace her Ladyship, the problem is that there was no attempt to balance the gay couple's right A against the hotelier's right B. The latter was simply brushed aside. That Strasbourg requires 'very weighty reasons' to justify discrimination on grounds of sexual orientation only dictates the stringency of judicial review; it cannot eschew the need for the Court to undertake the (admittedly) difficult task of assessing the respective restrictions of the competing rights by reference to the double proportionality test. An entire layer of the Court's analysis is absent. It should be stressed that applying the double proportionality test on the facts of the case would not have downplayed in any way the significance which Lady Hale so admirably placed on sexual and antidiscrimination rights. Rather, the main benefit of the test is to allow the Court to account for its conclusions via structured and rigorous judicial reasoning, and to ensure the rights of both parties are treated sufficiently seriously. As it presently stands, her Ladyship's judgment leaves much to be desired, for one is left in the dark as to why the homosexual couple's sexual rights – important as they are – trumped the religious rights pleaded by the hoteliers, given that both set of rights are recognised as fundamental under the Convention, and are equally entitled to judicial protection under the proportionality test.

This error is compounded in the recent decision of *Lee v Ashers*.[43] The facts echo those in *Bull v Hall*, albeit with some important differences. A gay customer ordered a cake from a bakery to bear the message 'Support Gay Marriage'. The Christian bakers cancelled the man's order because of their religious beliefs and refunded his money. The Supreme Court held that there was no discrimination on the basis of either sexual orientation, religious belief or political opinion. Of interest to the present discussion is Lady Hale's view[44] that, by being required to produce that cake the bakery was being required to express a message with which they deeply disagreed, and this violated their a right *not* to express an opinion under art 10. The principal competing rights here were concerned with freedom of expression, although like *Bull v Hall* the rights concerning manifestation of religion and respect for sexual orientation were part of the context as well.[45] The customer wanted to express a message on a cake, and the baker wanted not to express that message or to be associated with it in any way. It exemplifies an *intra-right* conflict, since the tension is within the same fundamental right.[46] Collins' chief criticism is that Lady Hale provided no explanation as to why the bakery's right prevailed: 'It does not seem inevitable that the right not to express a political opinion should be regarded as more important than the right to express one'[47]. That the case was litigated

[40] Discrimination law cases where the defendant relies on fundamental rights inevitably raise the question of how to balance the competing rights: H Collins, 'Justice for Foxes: Fundamental Rights and Justification of Indirect Discrimination' in T Khaitan and H Collins (eds), *Foundations of Indirect Discrimination Law* (Hart Publishing 2018) 255.

[41] *Bull v Hall* (n 39) [52].

[42] *Ibid* [53].

[43] [2018] UKSC 49.

[44] *Ibid* [49]-[58].

[45] H Collins, 'A missing layer of the cake with the controversial icing' (UK Labour Law Blog, 4 March 2019) <https://uklabourlawblog.com/2019/03/04/a-missing-layer-of-the-cake-with-the-controversial-icing-hugh-collins/>.

[46] L Zucca, 'Conflicts of Fundamental Rights as Constitutional Dilemmas' in E Brems (n 18) 26.

[47] H Collins (n 45).

as concerning discrimination law is no answer, for under s 3 of the Human Rights Act 1998 ('HRA'), the courts are obliged to 'repair' Acts of Parliament by interpreting them so far as it is possible to do so in a way that eliminates any potential violation of a Convention right. As acknowledged by Lady Hale: '[the relevant antidiscrimination legislation] should not be read or given effect in such a way as to compel providers of goods, facilities and services to express a message with which they disagree, unless justification is shown for doing so'.[48] Protection of the customer's legitimate rights under the Convention would no doubt qualify as a proper justification. The question remains: why did the bakery's rights prevail?

Several answers are possible. First, it might be thought that, on closer examination, this is not a conflicting rights case at all. The right to freedom of expression does not extend to a right to compel others to express views on your behalf. The customer's free expression right was not accordingly engaged. The conflict is spurious rather than genuine: the question of the content of a particular right should be kept separate from the question of conflicting rights (which concerns the situation in which a right makes something permissible while a competing right makes it impermissible).[49] However, this is not an entirely satisfactory answer, for the entire premise of private law contains a strong presumption that the rights and duties should be enjoyed universally and equally (this may be contrasted with discrimination law, which carefully identifies certain duty-bearers through legislation).[50] It is arguable that a bakery who offers to sell cakes to customers bearing messages of their choice should be considered a duty-bearer in respect of fundamental rights, and like the state, should be under a positive obligation to give effect to another's free expression right.[51] A second answer lies in the fact that the customer was able to obtain his cake from another bakery without difficulty. However, the availability of an alternative source cannot *ipso facto* tilt the balance in favour of the bakery, the same way that the availability of the opportunity for an employee to obtain alternative employment cannot be deemed in itself as a sufficient reason to give priority to the rights of the employer, should the employer be found to have committed religious discrimination against the employee.[52]

The third and best explanation is that, applying the double proportionality test, the customer's rights must yield to those of the bakery. One should not forget that the bakery could also claim the protection of art 9, the right to manifest their religious belief, on top of the freedom of expression. Applying the proportionality test in relation to the customer's free expression right (right A), the interference with that right was absolutely necessary to uphold the bakery's religious right (right B1), given the directors' strong and sincere disagreement with the relevant message with conflicted with their Christian beliefs; there was no less restrictive alternative on the table. The same goes for the bakery's freedom not to express a political view (right B2): there was simply no other way of upholding that right apart from permitting the bakery to refuse to bake the cake bearing the message they found objectionable. In contrast, applying the proportionality test – particularly stage 3, which considers the availability of less intrusive *alternative* means of substantiating the legitimate aim (of upholding the customer's right A) – in

[48] *Lee v Ashers* (n 43) [56].

[49] L Zucca (n 46) 25.

[50] See the Equality Act 2010 (UK), which covers the supply of goods and services, the supply of accommodation, employment, education and associations.

[51] H Collins, 'The Challenges Presented by Fundamental Rights to Private Law' in K Barker, K Fairweather, R Grantham (eds), *Private Law in the 21st Century* (Hart Publishing 2017) 213.

[52] *Eweida v United Kingdom* Applications Nos 48420/10, 59842/10, 51671/10 and 36516/10, Merits and Just Satisfaction, 27 May 2013.

relation to the bakery's rights, it is not the case that the customer's right A may *only* be upheld by preventing the bakery from exercising the full extent of their rights. No explanation was proffered by the customer as to why it was necessary to compel *this* particular bakery to voice his views on gay marriage. It would be different if, say, the bakery were the only one in the vicinity which offered the desired service of printing messages on cakes. It is only in this light that the availability of alternative sources should be material: coupled with the need to protect the bakery's rights B1 and B2, the fact that the customer had access to other bakeries who were more than happy to take his order meant that the interference with his right A was justified and proportionate.

Although my concern about balancing the free expressions might appear to be a minor point, it is in fact directly related to the legality of the bakery's behavior towards the customer. As Collins observes, the Supreme Court was in effect carving out a partial justification defence for direct discrimination cases in allowing the bakery to justify their conduct by relying on their own opposing beliefs, thereby avoiding liability for discrimination.[53] Whether the conduct could be justified on the facts should have been determined by a thorough application of the double proportionality test in relation to the competing free expression rights. To be sure, it would be different if, say, a Christian bakery rejected a heterosexual gay rights activist's order for a birthday cake simply because the former found the latter's support for gay marriage disagreeable. This would doubtlessly constitute unlawful discrimination on the basis of the customer's political beliefs, and there would be no question of competing free expression rights to resolve. Instead, the conflict, if any, would be between the customer's right against discrimination and the bakery's religious rights. And it would be most difficult, to say the least, for the bakery to justify their conduct as a legitimate manifestation of their beliefs.[54] But the Supreme Court did not go so far.[55] Lady Hale merely stated that this protection against discrimination on the grounds of political beliefs did not extend to compelling providers of goods, facilities and services to express a message with which they disagree, by reason of the need to uphold the fundamental freedom not to express an opinion. This conclusion accords with a proper balancing of the competing free expression rights engaged.

Both cases discussed thus far are examples of competing rights in the contractual context, concerning in particular the freedom to determine one's contractual partner and the subject matter of a contract (both of which are considered fundamental aspects of the freedom of contract). There is equally a possibility for rights to contradict one another in property law. A case in point is *McDonald v McDonald*.[56] The dispute concerned an action for possession against a tenant by receivers appointed by a mortgage lender. The tenant had severe psychological problems such that it would be extremely disruptive to her well-being for her to lose her home. Her parents acquired the freehold to the property with the help of a mortgage and fell into arrears. The tenant argued that whilst s 21(4) of the Housing Act 1988 allowed the receiver to obtain an order for possession in these circumstances, the interference with her art 8 right would thereby be disproportionate. The Supreme Court rejected this, and found that there was no

[53] H Collins (n 45); cf *R v Governing Body of JFS* [2010] 2 AC 728 [57].

[54] This is because the Christian bakery would have the uphill task of demonstrating a 'sufficiently close and direct nexus' between (i) refusing to produce a birthday cake for an activist for gay rights and (ii) their religious beliefs, failing which their right to manifest their religious beliefs under art 9 would not even be engaged.

[55] Lady Hale specifically distinguished between refusing to produce a cake conveying a particular message, for any customer who wants such a cake, and refusing to produce a cake for the particular customer who wants it because of that customer's characteristics: [55] and [62].

[56] [2016] UKSC 28. The decision was later upheld by Strasbourg, see *FJM v United Kingdom*, Application No 76202/16, Decision, 29 November 2018.

need to consider the proportionality of the order for possession. It is hard to see how this may be squared with the Supreme Court's earlier decision in *Manchester CC v Pinnock*,[57] where Lord Neuberger came to the opposite conclusion on analogous facts; the only discernable distinction is that *Pinnock* related to a public authority landlord, whereas *McDonald* concerned a private receiver. In both cases, however, there was a problem of competing rights which fell to be resolved by the Supreme Court. In *Pinnock*, Lord Neuberger's solution was to apply a single proportionality test in relation to the occupier's art 8 right, although his Lordship stressed two factors would weigh heavily in the authority's favour: first, the existence of the local authority's 'unencumbered property rights'; and second, the authority's 'right – indeed the obligation . . . to decide who should occupy its residential property'.[58] His Lordship added that in virtually every case, therefore, repossession by the local authority would be proportionate.

The difficulty with this guidance is that the proportionality test, as articulated by Lord Neuberger, is an unfamiliar one. The balancing is pre-weighted against the success of an art 8 defence advanced by an occupier, even if the occupier's defence is seriously arguable.[59] In fact, Lord Neuberger's concern is more accurately reflected by the recognition that competing rights would be engaged in such cases, militating against the occupier's art 8 right. Apart from the local authority's unencumbered property rights, his Lordship was also alive to the fact that other citizens might be affected by a failure on the part of the authority to exercise its obligation to decide who should occupy its residential property – that is, the fact that others eligible for public housing were just as entitled to protection of their right to a home under art 8. His Lordship was hence entirely correct in observing that in virtually every case, therefore, repossession by the local authority would be proportionate, but that is only by reason of the necessity to uphold the competing rights under the parallel analysis. Notably, this shows that in both public and private law cases there is a prospect for rights to conflict with one another, and for the double proportionality test to play an important role (*Pinnock* being a public law case).[60]

Returning to *McDonald*, there is no principled reason why the same double proportionality approach should not be applied to resolve the conflict between occupiers' art 8 right, and the art 1 Protocol 1 rights of private sector landlords (right to possession). If anything, the balancing exercise should be more straightforward given that, of the two factors mentioned by Lord Neuberger in *Pinnock*, only the first factor would bite in the private context. Indeed, one could take this further and contend that no balancing exercise is necessary at all by virtue of the 'inherent limitation' analysis. Lees opines that: '[S]ince the landlord's rights were always subject to the human rights of the tenant, and all art 1 Protocol 1 protects is pre-existing legal rights (as opposed to a factual state of affairs as does art.8), in fact there is no breach of art 1 Protocol 1 by giving effect to a tenant's art.8 right . . . '.[61] Whilst intellectually forceful, it might be considered to go too far in the opposite direction, particularly when one considers the mixed reception of the 'inherent limitation' argument in both Strasbourg and English jurisprudence.[62] The preferable view is that, in cases like *McDonald*, what is called for is a careful balancing

[57] [2010] UKSC 45; [2011] 2 AC 104.

[58] *Ibid* [54].

[59] R Walsh, 'Stability and predictability in English property law – the impact of article 8 of the European Convention on Human Rights reassessed' (2015) LQR 585, 599.

[60] The Hon Mr Justice Andrew Cheung PJ, 'Conflict of fundamental rights and the double proportionality test' (2019) 49 Hong Kong Law Journal 835.

[61] E Lees, 'Article 8, proportionality and horizontal effect' (2017) 133 LQR 31, 35.

[62] *Wilson v First County Trust Ltd (No 2)* [2003] UKHL 40, [2004] 1 AC 816; *J A Pye (Oxford) Ltd v UK* Application No 44302/02, Merits and Just Satisfaction, 30 August 2007.

exercise between the competing property and home rights through the lens of the double proportionality test. This need not lead to a 'wholly unpredictable and potentially very damaging' impact on the private rental sector.[63] As the Supreme Court in *Pinnock* recognised in relation to public housing, a summary judgment could be delivered in the majority of cases and that the situations where a proportionality assessment would have to be carried out would be quite rare. The same surely applies in the private context: the court could summarily make the possession order in the vast majority of cases, while leaving a margin for those few cases, involving genuine and arguable art 8 defences, where a double proportionality inquiry must be undertaken.[64] This is critical to ensure that the occupier's art 8 rights are not unduly disregarded and that a fair outcome is reached after a measured balancing process, with justice being done and seen to be done.

Taking a step back, it is evident that my argument on the merits of the double proportionality test is predicated on the assumption that judges would be prepared to get their hands dirty, so to speak, and challenge the way Parliament has reconciled the competing rights if necessary. There would be little point in devising a test to balance competing rights if the courts are not prepared to engage in the balancing process in the first place. In this regard, there is a worrying tendency on the part of the Supreme Court in such cases to rely on the premise that Parliament has struck a balance between competing rights to justify judicial non-intervention. For instance, commenting on the Supreme Court's rejection in *Bull v Hall* of the hoteliers' argument that their policy constituted a manifestation of their religious belief, Lady Hale explained extra-judicially that Parliament had not provided for such a defence in the Equality Act 2010 (which on the face of it simply prohibited discrimination against the customer on grounds of sexual orientation).[65] Again, in *McDonald*, the central plank of the Supreme Court's reasoning is that where the rights and obligations of private parties are determined by statute (namely the Housing Act 1988), it would be unfitting for the court to upset that balance. The Court specifically defended its deference to Parliament's stance by distinguishing cases concerning breach of confidence.[66]

This reasoning is unsustainable. The potentially relevant statutory provisions must, on any view, include the HRA, particularly s 3 and s 4.[67] As stated above, s 3 obliges the courts to interpret both primary and subordinate legislation so that their provisions are compatible with Convention rights; otherwise, s 4 permits the court to make a formal declaration of incompatibility. There is simply no basis to suppose the balance struck by Parliament between the rights of the hotelier and the customer under the Equality Act 2010, or the occupier and the landlord under the Housing Act 1988, should be considered conclusive. If taken to its logical conclusion, such an assumption would drain the judicial obligations under ss 3 and 4 of any meaningful content. The two subsections are intended to bite precisely when Parliament has legislated in a manner which affects the Convention rights of the parties: to assert, in such cases, that any restriction of those rights represent the considered view of Parliament, and that courts should for that reason alone refrain from intervening, flies in the face of what the HRA demands.

[63] *McDonald* (n 56) 43.

[64] JCB Sánchez, '*F.J.M. v. the United Kingdom*: Judicial review of the proportionality of an eviction in private rental housing' (Strasbourg Observers, January 15 2019) <https://strasbourgobservers.com/2019/01/15/f-j-m-v-the-united-kingdom-judicial-review-of-the-proportionality-of-an-eviction-in-private-rental-housing/>.

[65] Lady Hale, 'Are We a Christian Country? Religious Freedom and the Law' (Oxfordshire High Sheriff's Lecture 2014, 14 October 2014) < https://www.supremecourt.uk/docs/speech-141014.pdf> accessed 3 Aug 2019.

[66] *McDonald* (n 56) 46.

[67] E Lees, 'Article 8, proportionality and horizontal effect' (n 61) 35.

CONCLUSION

The 'constitutionalisation' of private law renders conflicting rights in private law claims increasingly frequent in this era of balancing. However, this process of 'constitutionalisation' should not be taken to support a wholesale incorporation of human rights discourses into private law, nor should it destroy the division between public and private law. It is necessary to translate public law conceptions of rights into a form and content suitable for reasoning in private law; the two parts of the law, private law and public law, ought to be normatively compatible, but those principles and rights cannot be expressed and articulated in the same way.[68]

The proposal in this article is, in the last analysis, a modest one: it is no more than a logical extension of the use of the single proportionality approach in public law claims, where only one party can invoke fundamental rights. For the same test to make sense in private claims, where both parties may raise arguments based on rights, it must follow that the double proportionality test has to be applied. Most fundamentally, the double proportionality enables judges to give adequate attention to the rights of all involved. Its adoption is imperative to achieve justice and fairness for both parties. No party should get a 'head start', or a presumption in his or her favour, in a conflicting rights claim. To take rights seriously means that courts have to adopt a procedure which gives due respect and concern to all competing rights engaged, and this must hold true regardless of the context of the private law claim.

[68] H Collins, 'Utility and Rights in Common Law Reasoning: Rebalancing Private Law Through Constitutionalization' (2007) 30 Dalhousie Law Journal 1.