3 Benefits, Not Consent: The Legitimacy of Institutions

> The state exists for the sake of man, not man for the sake of the state.
> — Verfassungsausschuss von Herrenchiemsee,
> *Entwurf eines Grundgesetzes* (1948: 61).[123]

3.1 Introduction

Normative phenomena such as rights and duties derive their existence and bindingness from institutions. What remains an open question, however, is whether it can be justified that people have these rights and duties. People who participate in an institution such as a political regime do not only want to know what rights and duties they *have* within this institution. They also want to know whether it is *justified* that there is such an institution which confers certain rights and duties to them and others. In the particular case of the state, we want to know as citizens and residents how it can be justified that rulers have the right to rule us and that we have the duty to obey them. This question refers to the *legitimacy* of institutions, i.e. the justification of their existence. The present chapter aims to provide an account of institutional legitimacy which answers to people's question for a justification of their own institutional duties in terms of the costs and benefits they individually obtain from participating in an institution.

Imagine a housewife in 1960s Germany. She is considering taking up a job as a bank clerk, the profession she trained for prior to getting married. Her former employer has expressed interest in getting her back on the team. Before they can come to an arrangement, however, she needs to consult her husband. He is not amused. "Who will cook my dinner, take care of the children, and dust the furniture if you work in a bank? Darling, your place is in the home. I won't grant you permission to engage in paid employment."

Let us assume that the relationship between the husband and wife is one of marriage in all relevant aspects. By rendering their signatures on a document, following a predetermined procedure in the registrar's office,

123 Own translation. In the German original: "Der Staat ist um des Menschen willen da, nicht der Mensch um des Staates willen."

they legally entered into wedlock. Additionally, a priest married them in a ceremony at the local church. Both wear a ring as a symbol of matrimony and share a surname. The husband provides financially for his wife and children and enjoys tax benefits in return. All these formal and informal social practices are constitutive for the institution of marriage at this time and place (although only the part at the registrar's office is legally required). Among these social practices are also certain legal rights accruing to the husband, e.g., to determine the family's place of residence or to bar his wife from working outside the home. Thus, she rings up her former employer and declines the offer.

According to the account of institutions I set out in the previous chapter, the husband's authority is real, creating binding obligations, because it is an institutional fact. That does not tell us anything, however, as to whether his authority, or the institution of marriage in general, is also *justified* to the wife.

Note that on the account I have so far advanced and defended, legitimate authority is not equivalent to de jure, i.e. binding authority.[124] De jure authority, that is the meta-right to create binding rights and obligations, can exist regardless of being justified, or even held to be justified, to those subjected to it. It may exist merely by virtue of its acceptance being required by an institution's rules of the game. In the previous chapter, I argued against the claim that governments lack de jure authority, which is put forward by philosophical anarchists. My ontological point was that philosophical anarchists conflate de jure political authority with a *moral* power-right, i.e. the right to create moral rights and obligations. Instead, taking a legal positivist position, I argued that political authority exists as a *legal* power-right that is different from brute power if and only if people in the state abide by the rule of recognition.

I did not, however, challenge the moral concern underlying philosophical anarchism, namely that there is something problematic with a government wielding political authority per se, even if we have reasons to comply with the laws it enacts. What I identified as the normative problem of political authority is exactly that rulers may derive real authority from a regime which is not legitimate, i.e. justified to exist. Likewise, the husband's authority derives from an institution the existence of which stands in need of justification.

124 This is in contrast to a typical usage in the literature which I described in 2.2.3.

The question of the legitimacy of institutions is indeed an important one, also and in particular from a legal positivist perspective. This is because legal positivism implies that people can have legal rights and duties, even though they *should* not have them from an evaluative standpoint. As set out in Section 2.3.3, institutions are binding as the rules of the game, not *qua* legitimacy. If you want to play the institutional game, you can only do so by abiding by the rules of the game. The rules are constitutive of the game; so playing the game is by definition abiding by its rules. Even outrageous social practices can be prescribed by binding requirements if they form part of existing institutions. This is why it is so important not only to know whether someone has a particular duty, but also whether the institution that entails this duty is legitimate. If an institution has been identified as illegitimate, this serves as a strong foundation for criticising it. Moreover, from a more practical perspective, the question of legitimacy also matters as a benchmark for abolishing or reforming existing institutions and for creating new ones.

To begin with, we first need to determine what exactly must be shown to be legitimate. Is it an institution such as political regimes or marriage per se, i.e., as institutional types? Or is it a particular institution such as the Federal Republic of Germany or marriage in our unhappy 1960s housewife example? The latter would be examples of institutional tokens. There are countless possible tokens of an institution. For instance, in the 2020s, the institution of marriage is still in place in Germany. Yet the social practices constituting this current institution differ in important respects from those observed during the mid of the 20th century. Among other things, adultery is no longer a criminal offense, whereas marital rape has become so. Marriage is also no longer exclusively a union between a man and a woman but open to adult couples of any gender. Moreover, none of the partners has the unilateral right to decide about the place of residence or occupation of their partner. There are many more manifestations of marriage as an institutional type across states, nations, cultures and eras.[125]

Given that differences among instantiations can be far-reaching, what is it that all tokens of the same institutional type have in common? Following Guala and Hindriks (2020), I individuate institutional types by their function. All institutions serve a certain etiological function as a *raison*

125 For instance, marriage-tokens may or may not involve practices such as divorce, paying a dowry, or polygamy. Also, tokens differ with respect to who is eligible to be married, e.g. only adults, heterosexual couples, or people from the same religious community.

d'être for at least some of their participants. Otherwise, they would not have come into existence. This function is coordinative and/or cooperative, depending on whether the social practices constituting the institution involve conventions, norms, or both.

The central function of marriage is to establish a particular relation of kinship which is exclusively institutional and does not originate in birth (aptly captured in the English language by the term "in-law" for family relations created through marriage). This main function is highly general and does neither entail nor exclude additional functions such as the raising of children. As marriage is an institutional relationship among sexual partners, offspring may often be involved. Yet elderly, infertile, and (in recent times) homosexual partners may also get married. Childrearing is therefore not the main function of marriage, although an important implication of marriage is that children born in wedlock are officially related to both partners and their respective families.

Note that serving an etiological function alone does not imply that an institution is beneficial for its participants. An institution may in fact make all participants worse off than they would be in in its absence. This is the case if the usefulness of its function is outweighed by the harm it causes, e.g. in the case of a hazardous dare performed as a rite of passage. There is a coordinative effect among the members of a peer group, but this could have also been achieved, for instance, by wearing the same kind of clothes.[126] With a sufficiently dangerous dare, the harm to the participants' health is far more substantial than the coordinative benefit achieved.

Moreover, an institution's coordinative or cooperative benefits need not accrue to each participating individual. Institutions may be lopsided, providing benefits exclusively to some of their participants while only imposing costs on others. Think of a caste system which excludes certain groups of people from particular occupations and from social power while granting access to others. This institution serves a function, but the function only benefits the upper castes.

Even though institutions may harm some (or even all) participants compared to a situation in which they are absent, such institutions can be stable. Individuals participate in an institution if and only if they prefer participation over non-participation. Yet this may also be the case if disobedience is judged to have even more adverse consequences than compliance with harmful rules, which is arguably not sufficient for an institution's legit-

126 A rite of passage is arguably a Hi-Lo coordination game (see 2.4.2), where the low equilibrium may actually yield negative benefits.

imacy. A participant who incurs net costs from the institution's existence will only concede that she has a reason to comply with the institution's rules *given its existence*. She will deny, however, that the existence of the institution as such is justified to her. Since each participant may equally call the legitimacy of an institution into question on these grounds, I will endorse a principle of legitimacy that states that an institution is legitimate if and only if it creates nonnegative benefits for *all* individuals incurring institutional burdens, irrespective of whether they choose to participate or not.

In principle, questions of legitimacy can be tackled at the level of both institutional tokens and types. It is fruitless, however, to consider the legitimacy of an institutional token if the institution has been found to be illegitimate as a type. A practice such as racism is arguably illegitimate in every instance because it imposes burdens upon a group of people without compensating them by means of benefits. This is the very function of racism. Such a function cannot be justified to all individuals who incur burdens from the existence of racism within a society.

The case of marriage is different: establishing an institutional kinship relation among sexual partners is a function which is not generally illegitimate. It may indeed create benefits for all parties involved. Some tokens of this institutional type, however, may pose grave issues of legitimacy. This is the case for forced marriage, particularly involving children. Very lopsided, that is patriarchal, tokens of marriage can also leave women worse off than they would be without any instantiation of marriage in place at all. Patriarchy is arguably also a practice that subjugates a set of people without compensating them for their burdens and therefore illegitimate at the type-level. Yet marriage itself can also take forms that avoid patriarchal patterns. Therefore, the institution is not by itself illegitimate, but only in some instantiations at the token-level.

Whether a token of marriage is legitimate is notably not determined by the fact that partners *consent* to get married. People may consent to marry because their parents threaten to put them into a monastery if they refuse, because they are pregnant and do not know how to support the child on their own, or because they need a residence permit in a country. Yet under such circumstances, taking a vow only indicates that the outside options are worse than participating in marriage, not that the existence of the institutional token itself can be justified to both partners. The practice of taking a vow is part of the institutional token of marriage, but it cannot justify it because it is only a formality and may not be a free choice. Similarly, an

oath of citizenship may be part of a regime's social practices. This practice, however, does not at all imply that the regime makes all of its subjects better off than they would be in some kind of state of nature. Consent need not track an institution's function of providing coordinative and/or cooperative benefits to participants. And not only may people consent under pressure. They may also deny their consent to an existing institution where everyone yields net benefits, simply because they want higher benefits for themselves.

If we are interested in whether an institutional token is justified or not, what matters is whether individuals consent to its *creation*, when no alternative token of the institutional type exists. For an already existent institutional token such as marriage in Germany in the 1960s, this question can only be raised hypothetically. Thus, we have to ask whether people who incur burdens from the institution's existence *would have* consented to its creation. These people are not necessarily only those who currently participate in it. An unmarried mother who faces social and legal stigmatisation may also incur costs from the existence of a particular token of marriage. Since she does not realise net benefits, this token is not justified to her. What is important is not whether she participates but that she would not face this burden if there was no instantiation of marriage at all. Other tokens may not come with such costs for non-participants and may therefore be justified to them.

In this chapter, I will proceed as follows: In Section 3.2, I will introduce my functional conception of institutional legitimacy, demarcating it against the notion that by participating in an institution, people acknowledge its legitimacy. Building on these elaborations, I will formulate a functional and individualist principle of legitimacy. In Section 3.3, I will set out how the functional conception of legitimacy can be illustrated by the thought experiment of a social contract, locating my approach in the contractarian tradition of the social contract literature. Section 3.4 discusses the merits of relying on hypothetical consent to a social contract in a counterfactual state of nature, compared to the criterion of actual consent. I will argue that hypothetical consent captures the fairness of existing institutional requirements, models voluntariness, which is hard to achieve under real-life conditions, and has practical implications for dealing with existing institutions based on their legitimacy. Section 3.5 provides a short summary.

3.2 Justifying Institutional Burdens to Individuals

3.2.1 *A Functional Conception of Legitimacy*

Institutions entail requirements for those who play by the rules of the game. These requirements can take the form of duties and obligations, or of more general behavioural prescriptions such as donning suit and tie in the case of a dress code. Compliance with institutional requirements imposes burdens upon participants which they would not incur if they had not chosen to play by the rules of the game. For instance, those who remain unemployed are free from the requirement to follow the orders of a boss, which they may dislike. Insofar as these requirements are burdensome, however, people may ask how the requirements can be justified to them. This question is particularly relevant for institutions such as political regimes or traditional marriage where some people are required to yield to the authority of others. Yet it is in no way restricted to these cases. People may also ask for the justification of a universal convention such as shaking hands, e.g. because they mind the hygienic implications of touching other people's hands.

In the following, I will use the concept of *legitimacy* as an evaluative term to refer to the justification of institutions.[127] As a modification to Rawls (1971, 3), I hold that legitimacy (rather than justice) is "the first virtue of social institutions."[128] Whereas the concept of legitimacy denotes

127 This means that I understand political legitimacy as referring to the legitimacy of political *institutions*. These are most notably political regimes (see Chapter 4) but also subordinate constitutional institutions (see Chapter 5). I do not, however, understand political legitimacy as a feature of political *decisions*. Peter (2023, 13–14), who focuses on the legitimacy of decisions in contrast to decision-making bodies, believes that this framing is merely a methodological question. It does, however, have substantial implications. The functional account of legitimacy which I am developing in this section is only applicable to institutions, not to singular decisions, because only institutions have functions. Her focus on decisions, moreover, helps explain why Peter (2023, 91–101) suggests a conception of political legitimacy which has an epistemic and a voluntarist ground, as decisions are influenced both by an agent's will and her knowledge. (It also explains why she gives lexical priority to the epistemic component, since we typically want to make the right decisions.)

128 See also Larmore (2020, 83) who emphasizes that the state's primary function is not to establish social justice but to provide order. For Kukathas (2003, 260), too, the fundamental question of political philosophy is the question of political legitimacy, rather than the question whether a society is just. Such a conception of legitimacy is in contrast to Brinkmann (2024) who understands legitimacy in terms of justice. His notion of justice builds upon the primary values of welfare and dignity as well as other secondary values, such as democracy, and entails moral rights of individuals.

the justifiability of an institution's existence, *justice* is an evaluative term for distributions, i.e. the distributive dimension of institutions, actions and outcomes. Thus, an institution may be legitimate even if it is neutral in terms of justice, simply because it does not deal with distributions. For instance, traffic regulations may be legitimate, i.e. a justified institution, even though they cannot actually be described as just. The term legitimacy can be used with respect to all sorts of institutions. *Political legitimacy* refers to the legitimacy of political regimes in particular.

After defining the *concept* of legitimacy, I will now turn to the question which *conception* of legitimacy to apply in order to judge the justification of institutions. First of all, I take it that institutions must be justified *to all individuals incurring burdens* to be justified at all, i.e. legitimate. This is a *normatively individualistic* position. Normative individualism is the contention that the relevant unit at which a justification is to be directed is the individual.[129] The reason for adopting normative individualism as the normative basis for an account of institutional legitimacy is formulated concisely by Føllesdal (1998, 199):

> The only ultimate bearers of value are individual human beings. Thus arguments regarding the legitimacy of social institutions (including associations and nation-states) must be made in terms of how they affect the interests of all affected parties.

Insofar as all individual participants face burdens from institutional requirements, they all have a claim to ask for a justification of these burdens. And since all participants may ask for an institution's legitimacy, a justification must be given to all of them.[130] This justification, moreover, must be one that each individual can accept. Otherwise, she can always claim that the institution is not justified *to her*.

Rawls (1971, 363) himself also ascribes legitimacy to a regime in virtue of its justice, although he does not formulate a conception of legitimacy.

129 Vanberg (2004, 154) defines normative individualism in the context of evaluating political regimes as "the assumption that the desirability and legitimacy of constitutional arrangements is to be judged in terms of the preferences of, and the voluntary agreement among, the individuals who live under (or are affected by) the arrangements."

130 A similar point is made by Gaus (2011, 268) who holds that the public to which a rule has to be justified is defined by the participants of the social practice that the rule regulates.

Moreover, the conception of institutional legitimacy I want to put forward in the following draws upon the function of institutions.[131] In the most general formulation, the function of institutions is the creation of coordinative and/or cooperative benefits (see 2.4.2). Benefits and costs are understood here in a very broad sense as everything that increases or diminishes a person's utility. The terms are thus not restricted to monetary values—and also not to Rawlsian primary goods.[132] If there were no benefits to be gained, institutions would not have evolved or been created.[133] It is thus the benefits of coordination and cooperation which people ultimately care about in institutions. We may talk of institutions such as marriage as if they had a value on their own. Yet ultimately, the value of marriage is that it enables partners to enter into a committed partnership which is recognised by the government, enabling them to coordinate and cooperate with each other, as well as with other members of society.

A great merit of the functional approach to institutional legitimacy is that it does without demanding presuppositions. The justification of institutions cannot itself rely upon institutions, as this would be circular. This does not only rule out the formal institutions of a legal order, but also institutional phenomena from the sphere of social morality such as moral rights. Social morality provides a helpful institutional framework to justify our behaviour in everyday life. Yet insofar as it is also an institution, it also stands in need

131 Common alternative conceptions of (political) legitimacy are based upon hypothetical or actual consent, the principle of fair play, or the "normal justification thesis" suggested by Raz (1990, 129–130). For an introductory overview of common theories of legitimate practical authority, see Wendt (2018a). A very different take on the matter, moreover, is provided by Fossen (2024) who understands political legitimacy as an existential predicament and discusses what it means to make judgments of legitimacy rather than offering criteria for a legitimate regime.

132 Rawls (1971, 62) defines primary goods as material and immaterial goods virtually everybody strives for. There are both social primary goods, such as rights, income, or self-respect, and natural primary goods, such as health or intelligence, which withstand the forms of redistribution available for social goods.

133 Apart from an institution's *etiological* function, that is the benefits which explain its existence, Hindriks and Guala (2021, 2036) also identify a *teleological* function, i.e. its contribution to a certain value. Whereas the etiological function explains the existence of an institution, they claim, it can be evaluated by reference to its teleological function. I do not make this distinction between different types of functions since I conceptualise both the reason of their existence and the legitimacy of institutions in terms of benefits.

of a justification.[134] Moreover, the criterion for justifying institutions cannot itself rely upon institutions since that would be circular.

Consent, too, is an institution and thus stands in need of a justification. The option of binding ourselves by means of consent is only available for us because corresponding social practices are passed on over the generations and acquired in childhood (Pitkin 1966, 46–47). The function of consent is to enable individuals to waive their moral or legal rights, e.g. in the case of medical interventions, or to take on new obligations by entering into binding commitments, e.g. in the case of marriage. Justifying the bindingness of consent in a particular case by referring to an earlier instance of consent would beg the question why that instance of consent is justified. Ultimately, invoking earlier and earlier instances of consent would lead into an infinite regress.[135]

No such problem of circularity or an infinite regress arises for a justification which evaluates institutions in terms of the coordinative and/or cooperative benefits they provide for their participants and everyone else who incurs institutional burdens. Without invoking other institutions, such as altruistic norms prescribing selflessness, which would themselves need to be justified, the only justification of an institution's existence that a prudentially rational individual is going to accept is that it yields benefits to her.

I am not, however, formulating a *functionalist* conception of legitimacy. What I take to be functionalism is the position that an institution is justified by virtue of the fact that it serves or once served a function. Such a function, however, may just be to create privileges for a particular social class, at a cost to everyone else. That an institution *has* a function thus provides an explanation but not a justification of its existence. Since their function of creating benefits is the reason why institutions exist at all, a functionalist account of institutional legitimacy would need to classify all existing institution as justified, simply in virtue of their existence. In this way, the justification of institutions would be reduced to the social fact of its existence. It is, however, exactly because institutions exist that participants may ask for a justification of their institutional burdens. Merely pointing out to participants that an institution exists because it serves (or once

134 As Moehler (2018, 147–148) points out, that the rules of social morality originate in social evolution rather than political authority can only explain but not justify them.

135 Stemmer (2013, 3–5) makes the same point with respect to the institution of contractual agreement.

served) a function would beg the question why it should be justified to them.[136] It is important here not to commit the naturalistic fallacy and infer an *ought* from an *is* (see 2.4.1).

Institutions are tools, like knives, which may be more or less beneficial, or even harmful, for their participants. Pointing out that an institution has *some* function need not be a justification that satisfies all its participants. Individuals who incur costs from an institution's existence want to know that overall, the institution creates benefits *for them*. I therefore refer to an institution as *functional* if and only if the burden it imposes on individuals can be justified to *each of these individuals* with coordinative and/or cooperative benefits they receive in return.[137] According to this functional conception of legitimacy, a functional institution is legitimate, but it is not sufficient for functionality that an institution serves a function. Even overtly discriminatory and harmful institutions have a function; this is why they exist and persist. Yet their function is to create benefits for only some of their participants, while others face nothing else than burdens. The continued existence of an institutional token must therefore not be misinterpreted as a sign that this institution serves a function for everyone who incurs an institutional burden.

Discriminatory institutions need not be the product of malign intention. Although such institutions can of course be actively created, they may also emerge as the result of social evolution (see also Buchanan and Powell 2018, 253–254), and their beneficiaries need not even be aware of it. Examples for discriminatory institutions are patriarchy, caste systems, nobility, and racism. The function of these institutions is to create a higher social rank with a particular practical authority and social power for a defined subset of the population, e.g. men, members of high castes, nobles, or a particular

136 This position is also taken by Greene (2019, 214–215) who claims that "[w]hen we are in the domain of social practices, we cannot evaluate their legitimacy without first identifying an implicit claim about their purpose, their *raison d'être*. In these cases, I suggest, legitimacy depends on recognition by participants that this claim has been fulfilled."

137 Pettit (2023) also emphasizes that regimes must be functional. In contrast to the usage here, however, he employs the term to denote a regime's stability (by virtue of providing benefits of security to citizens), rather than its justification. As I will discuss in the next section, individuals can have incentives to participate in an institution even if it is not justified to them. Pettit's notion of functionality is thus not even a minimal criterion of legitimacy; it is not related to legitimacy at all. Nevertheless, Pettit (2023, 262–63) holds that in order to be functional in his sense, a state must guarantee some substantial and equal rights at least for the citizenry (which need not include all residents).

ethnic group. Those against whom this discrimination is directed will typically not give an affirmative answer when asked whether they benefit from the institution (even though they may, as a result of internalisation).

It is not uncommon for institutions to exhibit discriminatory characteristics. These are grounded in the salience of obvious asymmetries between players. Exploiting asymmetries may be mutually beneficial, insofar as it may help individuals to coordinate on a social practice when they have partially conflicting interests.[138] For instance, at a crossroads, everyone would like to go first, but more importantly, they want to avoid a crash. Attributing the right of way to vehicles coming from the right, which exploits an asymmetry between vehicles coming from different sides, achieves this coordinative function and thus yields coordinative benefits.

Agential features that break symmetry may, however, also consist in traits such as gender or ethnicity (Sugden 1986, 92–93). Such features are highly salient. Yet by using them to coordinate, they become institutionalised themselves as social categories, e.g. when the biological feature of sex forms the basis of the institution gender, or when the category "race" is constructed from external features such as skin colour. The emergence of such categories may then lead to forms of discrimination that lack coordinative value for those affected by it.

To understand the evolutionary origin of gender as a social category, it is helpful to turn to Cailin O'Connor's (2019) account. She considers gender as an evolved behavioural pattern, building upon but distinct from biological sex differences, which solves a population-wide complementary coordination problem (see 2.4.2).[139] The emergence of the social category gender with the types "woman" and "man" is capable to transform the complementary coordination problem of dividing household labour into a conditional correlative coordination game. That means that all individuals follow the same rule which conditions their behaviour by type, such as "step

138 Skyrms (1996, 66–79) argues that in iterated mixed-motives games, a *correlated equilibrium* in pure strategies becomes available by assigning strategies to players based on a salient feature that breaks symmetry. The advantage of such a correlated equilibrium is that, although introducing inequality, it yields higher average payoffs than playing mixed strategies. As Hindriks and Guala (2021, 2030) note, a correlating device such as a traffic light extends the set of possible strategies by conditioning behaviour.

139 O'Connor (2019, 98) notes that gender is particularly likely to emerge as a social category because the marker of biological sex is highly salient due to its reproductive role, but also because the population is evenly divided between males and females, and households typically consists of one adult member of each type.

forward if you are a woman and back if you are a man" in dancing. To make coordination even easier, types are emphasized by means of ostentatious signals. Individuals then use these type-signals to condition their behaviour in coordinative situations (O'Connor 2019, 38–43).

Whereas social categories are conducive to efficient coordination, they also invite discrimination because individuals cannot simply change types (O'Connor 2019, 53). What is more, once social categories such as gender exist, they also allow for unequal outcomes in distributive bargaining games where inequality does not serve a coordinative function such as the division of labour (O'Connor 2019, 107–11). Thus, a society may become permeated with sexist social practices which cannot be justified to women on the basis of any benefits they would gain.

For instance, informal institutions such as foot binding, female genital mutilation,[140] or honour killings[141] may lead to the mutilation and even murder of women. The burden which women suffer from these institutions is brute violence to which they are being passively subjected. In these cases, women who are killed and mutilated by their own families are victims, rather than participants in institutions. It is the relatives who participate in social practices of murdering and injuring their own daughters and sisters, to restore family honour or ensuring them a good match. A mutilated girl need not take the internal standpoint and acknowledge any institutional requirements but may still ask for a justification for the harm she suffers as a consequence of the institution's existence.

140 Mackie (1996) compares the historical practice of foot binding in China and female genital mutilation which is still practiced in parts of Africa. Both are similar insofar as they are or were informal social practices, performed by women to restrict other women's (mostly their own daughters') sexuality in order to ensure prospective husbands of the paternity of the woman's future children. According to Mackie, the background is that in unequal, polygynous societies, men have difficulties to control the fidelity of their wives. Families will subject their daughters to such damaging practices as foot binding and female genital mutilation in the hope to marry them off to the men with the highest status.

141 Handfield and Thrasher (2019) discuss the emergence of honour codes. They argue that "norms of purification," an extreme case being so-called honour killings, serve the function of a costly signal. A family thereby indicates that even though one of the daughters behaved "dishonourably," the other children are still good candidates for marriage. Such a signal is economically and/or biologically important for the family (see also Thrasher (2018a)).

3.2.2 *The Participation Constraint*

Institutions may fail to be justified to non-participants who incur costs from their existence. It is important to note, however, that merely because a person participates in an institution and acknowledges institutional requirements, this does not entail that the institution is justified to her. Indeed, by choosing to participate, she obtains institutional benefits which would otherwise not be available to her, and she reveals that she values having these benefits more than the alternative. Individuals who participate in an institution thus have a preference for participation over their respective outside options.[142]

If the housewife from the 1960s asks her friend how it can be justified that her husband has the right to keep her at home, the friend might retort: "Since you wanted to get married, you now need to obey your husband. You get the benefits, so you also have to bear the costs. These are the rules of the game." Among these benefits is the fact that her husband is obligated to provide for her and their children. At the same time, however, she incurs costs in the form of institutional requirements that are also part of the "rules of the game." For instance, among the housewife's costs from marriage is the fact that she must obey her husband's authority. That cost may be quite substantial to her and only worth bearing because, in her society, the alternatives are even worse.

She might thus reply to her friend that she did not even want to get married. The reason she did so in the end was that unmarried women suffer a huge disadvantage, and even more so if they are mothers. Outside of marriage, in contrast, the housewife could have achieved the benefit of working as a bank clerk (although with few prospects of career advancement). A cost would have been that she could not have had children without incurring social stigma and legal as well as financial disadvantages for herself and her children (for instance, they could not have been their father's heirs). Since she wanted to escape her strict parents and to have children of her own, getting married was the best available alternative to her, even though it was by no means an alternative she liked.

142 Note that preferences differ from desires. As Gaus (2011, 311) points out, it is possible to prefer one bad option to another, while desiring none of them. Heath (2008, 23), moreover, stresses that desires can be in conflict with each other, e.g. for going to the cinema and staying home. A preference, in contrast, uniquely identifies what an individual likes best, all things considered, in a given situation.

To motivate participation in an institution, it is accordingly sufficient that the combined benefits and costs from acknowledging institutional requirements outweigh the combined benefits and costs from not participating. In technical terms, an individual *i* decides to play by the rules of an institutional game *x* if the institution meets a *participation constraint* for her. That is the case if the total utility U_i she can achieve from participation outweighs the total utility she could gain from not participating. The individual's utility U_i can be understood as the sum of the costs (i.e. institutional burdens) and the coordinative and/or cooperative benefits the individual *i* realises in each scenario. Formally, this relation can be expressed as

Participation Constraint (PC): *Ui* (participation in *x*) > *Ui* (no participation in *x*)

If PC would entail functionality, then every existing institution would be justified to all participating individuals by virtue of its continued existence. However, this is in conflict with the fact that individuals may continue to participate in an existing institution, thus perpetuating its existence, even though the existence of the institution serves no function for them.[143] This may occur insofar as the costs from non-participation are a consequence of the institution's existence, such as sanctions for non-compliance.[144] In this way, the utility from non-participation may be even lower than from participation, even though, in the absence of the institution, the individual would not benefit from its introduction. Since defiance of patriarchal (and other discriminatory) norms is punished by social ostracism, and in some countries even by formal sanctions, most women in patriarchal societies prefer to play by the rules of the game and to submit to men's authority. They can deny, however, that the existence of patriarchy is therefore justified to them, since they are worse off with patriarchy than they would be without it.

For an individual to accept an institutional token as justified to her, she must be better off given its existence than without it.[145] Thus, she must

143 Gaus (2011, 435) actually holds that informal norms which oppress women or ethnic groups are not capable of maintaining their status as norms. However, discriminatory institutions are not inherently unstable. To the contrary, once they exist, they may be hard to abolish because people have incentives to participate.

144 Lawless (2025, 1157) also observes that some social norms which exist because they benefit some, but not all, members of a society can persist because those who benefit are in the position to make deviance costly.

145 See also Gaus (2011, 237) whose notion of the "eligible set" contains those and only those rules which are pareto-superior to having no binding rules in these types of

yield *net* benefits from the token's existence. In other words, the sum of benefits and costs she obtains due to the existence of the institutional token must not be negative. It does not suffice that she yields benefits from participation compared to non-participation once the token is already in place. Rather, the baseline of comparison must be a situation where the token in question does not exist, nor any other token of the institutional type. Insofar as there already exists a token of the type in question, the situation of comparison must be a counterfactual one which abstracts from reality in this respect. If we think about introducing a new institution, in contrast, we can take the world as it is now as our baseline. In the case of marriage, the relevant baseline would be a counterfactual scenario where there is no formal and/or informal form of marriage. For political authority, the non-institutional outside option would be some kind of state of nature without formal institutions and authorised power.

Such a non-institutional baseline is required because the question is whether, from the perspective of the individual, this token serves the function of its respective type or not. If there was another token of the same type, her evaluation would depend on whether she can achieve more benefits than with this other token. If these benefits were high, she would reject many functional ones. For instance, women today would not approve of the introduction of a more traditional form of marriage because their own benefits would be lowered by such a measure. In the counterfactual situation, however, they might be in favour of it because they benefit from the possibility to create an institutional kinship relation to their sexual partner. This would mean that the institutional token serves a function for them, although their benefits could be higher with an alternative token.

If the existing token was very oppressive to the individual, however, she would even prefer a small reduction of costs without the prospect of net benefits. For instance, a woman might prefer a token of marriage where she is allowed to work without her husband's consent, although marital rape is not criminalised. In this case, the benefits from having the institutional token would not outweigh the burdens she might incur. Even though the new token is better for her than the old one, none is actually worthwhile for her to have at all.

---

situation. In contrast to the hypothetical contractarian approach followed here (see 3.3.2), however, Gaus takes a public reason approach which works with idealising assumptions concerning the individuals to whom a rule must be justified. I argue against idealisation in 4.4.3.

In either case, the individual's judgment would not tell us whether the token to be evaluated actually solves a problem of coordination and/or cooperation from her perspective. It merely contains the information how it fares compared to the existing institutional setup. Only the fact that individuals would prefer to have an institutional token compared to such a non-institutional scenario is indicative that she actually benefits. Taking its existence in real life as given, she may prefer to participate in an institutional token to not participating. But she may always challenge the claim that the token is legitimate based on the net costs she incurs from its existence.

Accordingly, the housewife could dispute her friend's assertion that the existence of marriage in 1960s Germany is justified to her, even though she participates in it. The benefits are enough to incentivise her to get married. Yet they need not be sufficient to justify to her that there should be this token of marriage in the first place. This is because the burdens of unmarried motherhood, which form part of the costs of non-participation, are a consequence of the fact that this particular institutional token of marriage is in place. That the benefits of getting (or remaining) married are higher than the alternative does therefore not entail that the existence of this token of marriage serves a function for her. It only means that now that the token is in place, it is worthwhile for her to play the game and abide by its rules, i.e. to get married and to recognise her duties as a wife.

This is somewhat similar in the political sphere. Given the existence of a regime, playing by the rules of the game and acknowledging the legal order as binding is usually more attractive than defiance. A benefit which people gain from acknowledging a legal order is the possibility to claim legal rights, e.g. to property. Those who do not recognise the rulers' authority and the law's bindingness, however, do not merely forego legal benefits. They are being threatened to comply with brute power when the executive gets hold of them. This prospect may be worse for them than a situation where no regime exists and thus no rulers wield authorised power.[146]

Just as an institution may be unjustified to people who participate in it, it can also be legitimate to impose costs on those who do not acknowledge institutional requirements, choosing not to participate in an institution.

146 Pettit (2023, 145–46) claims that for citizens to accept a sovereign's authority rather than yielding to his or her power, they must gain some benefits, e.g. of coordination, from the legal system. If they are merely afraid of the consequences of non-compliance, there is no acceptance, he holds. Yet for some people, the fact of avoiding sanctions may already be enough benefit to incentivise them to play by the rules of a regime-game, even though they will not consider it as justified.

This is because it is possible to benefit from an institution without acknowledging a duty to participate in it.[147] For instance, you may deny that you have a duty to assist an injured person, even if it comes at a low cost to yourself. This duty, however, is arguably justified because you benefit from the prospect of being helped and possibly saved when injured, while the costs to you are moderate (per definition).[148] So it is justified to convict you for failure to render assistance if you let a person die whom you could easily have saved. Similarly, a free rider on public transport may legitimately be fined if she benefited from the ride, although she may not recognise a duty to buy a ticket. Thus, an institution is not justified to individuals insofar as they participate in it and recognise institutional duties, but insofar as they benefit from the institution's existence.

3.2.3 *The Principle of Legitimacy*

So far, it has been established that a functional justification of an institutional token's existence to an individual must invoke the benefits she obtains due to its existence. Moreover, these must be net benefits compared to a counterfactual baseline scenario without any token of this institutional type in place. Combined with the individualistic requirement that, to be legitimate, an institutional token must be justified to each individual who may ask for a justification because she incurs institutional cost, this leads to

> *Principle of Legitimacy (PL)*: An institutional token is legitimate if and only if its existence does not impose positive costs on any individual, compared to a counterfactual situation without any tokens of the respective type.

Note that PL is a condition of Pareto indifference compared to the situation where no institutional token of the type in question exists. On the functional account, legitimacy is measured in terms of costs and benefits, but these are not aggregate benefits but the benefits of discrete individuals. Thus, functional legitimacy is not a matter of charging up the benefits of one

147 Among those who do not recognize the legal order are criminals who break primary law, terrorists who fight the constitution, and illegal migrants who cross the state's border without authorisation.

148 It is debatable, of course, how high costs may become before such a duty cannot be justified any more.

group against the costs of another.[149] That is to say, PL is not a utilitarian principle. This follows from the assumption of normative individualism. Even though functionality is a matter of costs and benefits, benefits must at least equal costs for each participating individual. The single individual is not impressed by the fact that an institution creates a high total amount of legitimacy, as long as she faces net costs. Thus, a social practice which benefits a large majority at the net expense of a small but oppressed group is as dysfunctional as one which oppresses a great number to the benefit of a narrow, privileged elite. As long as the institution is not redesigned to compensate those realising net costs for the existence of the institution, it is illegitimate.

PL is also not an egalitarian principle. Beyond the requirement that nobody must be worse off with an institution than without any token of this type, there is no specification how benefits are to be distributed among participants. Accordingly, battle-of-the-sexes conventions constitute functional social practices, even though one party achieves higher gains than the other. This is why traditional forms of marriage may indeed be legitimate, on the condition that women are still better off than they would be without any token of marriage as an institutional type at all. This may be doubted in the case of our housewife, since marital rape was not a criminal offense in Germany in the 1960s. Arguably, the fact that rape is not punishable as rape if it occurs within marriage makes women worse off than they would be without marriage. This would mean that such tokens of marriage are *dysfunctional*.[150] In the end, however, it is an empirical question how high individuals evaluate certain costs and benefits and how they weigh them against each other.

What also does not matter for functionality is whether a pareto-improvement to this institutional token is possible, i.e. if there is a way to change the social practice(s) such that all participants would be better off by means of saving opportunity costs.[151] Conventions which form suboptimal equilibria

149 In contrast, Hampton ([1997] 2018, 98–99) holds that political authority is justified if the moral costs it entails are smaller than the moral costs of having no authority, but she does not rule out an aggregation of costs across individuals. This is an important difference to my approach.

150 My use of the term "dysfunctional institutions" bears some similarity to how O'Hara and Ribstein (2009, 21) employ of the term "bad laws" for laws which impose net costs on parties subject to these laws.

151 But cf. Gaus (2011, 434–43) who, in the context of social morality, identifies three different types of "bad" rules: (1) unjustified self-enforcing equilibria, (2) unjustified

in Hi-Lo games, such as awkward dress codes, are therefore functional as well, as long as individuals are better off with them than without any coordination, independent of how they would benefit from more comfortable alternatives. A special case of dysfunctional informal institutions are those which are detrimental to *all* their participants. These include practices that are induced by peer pressure, such as substance abuse,[152] unprotected sexual intercourse, criminal conduct, or high-risk dares as passage rites. The desire to conform can induce individuals to engage in practices which, in total, do them more harm than good, even though they gain some benefits of coordination within a peer group.

The criterion of functionality is applicable both to institutions as sets of social practices and to individual social practices in isolation. In particular, it may be the case that an institutional token which is by and large functional can include some dysfunctional social practices. This is not uncommon. Consider again the running example of marriage in 1960s Germany. Even under the assumption that both men and women obtain net benefits from entering this legal kinship relation, it may still be the case that some of the social practices associated with the institution are harmful overall to women. Among these harmful social practices are arguably the husband's right to veto his wife's paid work outside the home and his sole right to determine the family's place of residence. Women would be better off without these social practices, even if it was the case that they benefited in total from the existence of marriage. In the same way, a legal order which includes a dysfunctional institutional token of marriage can still be legitimate as long as the legal order as such is functional. The more complex an institution, the harder it will be to avoid any dysfunctional social practices or subordinate institutions (see 5.4.4). However, even though they can be individually criticised as illegitimate, dysfunctional subordinate institutions do not necessarily impair the legitimacy of the institution itself.

Moreover, a functional institutional type may also have dysfunctional tokens. An institutional type is functional if and only if its function is one which does not entail net costs for any individuals. In other words, its function must be acceptable to all individuals who incur institutional burdens. On this account, marriage is a functional institutional type. The creation of an institutional kinship relation among sexual partners a such

equilibria kept up by punishment, and (3) non-optimal moral equilibria, i.e. rules that are not Pareto-optimal.

152 Pettit (2023, 42) also gives the example of drinking heavily in a peer group for a harmful rule.

is a function which can serve all participants in the institution without imposing costs on non-participants. Thus, the institution of marriage is not inherently unjustified, although it has been historically associated in many cultures with patriarchy.

Nevertheless, some tokens of marriage are clearly dysfunctional. Forced marriage, for instance, comes with net costs for the victims of such an oppressive institution. Moreover, the benefits of creating an institutional kinship relation among sexual partners can only arise among adults. All tokens of marriage involving children as spouses are therefore dysfunctional and unjustified. It may even happen to be the case that all tokens of a functional institutional type which have been realised as of now are dysfunctional and thus unjustified. For instance, all existing tokens of marriage may be too patriarchal to count as functional. Insofar as the function of marriage as a type is functional, however, it would be theoretically possible to create a functional token.

In contrast, there are also institutions which are unjustified at the level of types, such as slavery, apartheid, or patriarchy.[153] The whole function of such institutions is to oppress or downgrade a set of people who incur net costs from the existence of the institution. As dysfunctional institutional types are unjustified, any possible token of them is unjustified as well. Since it is the function of slavery to exploit the slaves' labour to the benefit of the masters, there is no instance of slavery that could be justified to slaves.

3.3 Legitimacy as Hypothetical Consent to a Social Contract

3.3.1 *The Notion of the Social Contract*

On the functional account of legitimacy introduced in Section 2 of this chapter, legitimacy is defined in terms of the costs and benefits that individuals face as a consequence of an institution's existence. This is in

153 The claim that institutional types may lack justification is challenged by Guala (2016, 199). He holds that normative evaluations can only be appropriate at the level of tokens whereas institutional types may only be described but not evaluated. Yet this claim comes at the huge cost of not being able to condemn institutions such as slavery as dysfunctional. With respect to slavery, Guala (2016, 5) is even committed to saying that it is at least slightly beneficial for slaves, claiming that the noninstitutional alternative would be genocide. This is an ad hoc assumption without empirical foundation. Moreover, it diminishes the suffering of slaves to claim that they benefited from the existence of the institution of slavery.

line with the idea that an institution's function, i.e. the reason for its existence, is to create coordinative and/or cooperative benefits. If we want to identify which institutions are legitimate and which are not on this account, however, we face a practical obstacle. Since costs and benefits are subjective evaluations of individual people, their values are not actually accessible from an outside perspective. We thus need to rely on an auxiliary device, namely the thought experiment of a *social contract*. The functional conception of institutional legitimacy can therefore be understood as a generalisation of hypothetical social contract theory.

To illustrate what counts as a legitimate constitution for a regime, hypothetical social contract theory uses the metaphor of the social contract which is unanimously ratified by all individuals in the state of nature. The state of nature is not a historical phase of human evolution.[154] Rather, it is a counterfactual situation where no state-token, and therefore also no regime, is in place. Since individuals would only accept the creation of a new institution which makes them at least as well off as they are without it, unanimous hypothetical consent to the social contract in the state of nature tracks functionality. It entails that no individual yields net costs from the regime's existence. The fact that the adoption of the social contract must be unanimous means that everyone has a veto to block a constitution which is not acceptable to them.[155]

Note that the role of the hypothetical social contract is *not* to explain why people are bound by the rules of a regime (or any other institution). As discussed in Chapter 1, institutional requirements are binding for those who participate in the institution and therefore need to abide by the rules of the game. The hypothetical social contract, in contrast, *illustrates legitimacy*, i.e. what it means that a regime is a Pareto-improvement compared to the state of nature. By agreeing to a social contract, individuals in the state of nature reveal that they would be at least as well off under the regime it defines as they are under their current circumstances. This is the difference to the participation constraint (see 3.2.2): Individuals do not only prefer to

154 This is not a new interpretation of the concept of the state of nature. Already Hobbes ([1651] 1996, 89–90), who conceptualises the state of nature as a state of war, notes that the state of war does not describe a phase in history. He claims that civil war can bring about the state of nature even where people used to live under a government. Hume ([1739] 1960, 493), moreover, stresses that the state of nature is "a mere philosophical fiction, which never had, and never cou'd have any reality."

155 Popper ([1945] 2013, 108–109) claims that hypothetical social contract theory captures the idea that the state is a means to the end of protecting weak individuals.

participate in an institution given its existence; they also prefer its existence to its non-existence.[156]

The thought experiment of the social contract can easily be adapted to all other types of institutions apart from political regimes, e.g. marriage. To that end, the state of nature merely has to be exchanged for the counterfactual scenario where no token of the respective institutional type is in place. What remains the same is that all individuals who incur costs from this token would need to consent to its introduction.

Hypothetical social contract theory may also be employed for evaluating the legitimacy of social moralities and their respective institutions and social practices.[157] Whereas the function of political regimes is to ensure peaceful coexistence within a state (see 4.2.1) as a political organisation, the function of social morality is to regulate human coexistence within moral communities.[158] These communities need not coincide with the population of a state. Moreover, they typically exist on different scales or levels.[159] Lower-level moralities can be exclusive and lay claim to regulating the lives of their members quite closely. For evolutionary reasons, such moralities are likely to require larger sacrifices from the individual, thus being more utilitarian than the morality of the wider society (Binmore 1994, 24). In the extreme case, people may even be required to sacrifice their lives for the community. This clearly makes the individual worse off than she would be in a fictitious pre-moral state where she is at least granted the possibility of

156 Lewis (2002 [1969]: 92) also notes that in the case in which a convention stabilizing a sovereign's rule exists but some or all individuals would prefer the state of nature to this status quo, the convention is *not* a social contract.

157 Stemmer (2013), for instance, adopts the *prohibition of oppression* (own translation) as a social contract criterion for the evaluation of social-moral norms. The prohibition of oppression requires that moral norms must serve the interests of all members of the moral community to which they apply. In a similar vein, Hart ([1961] 2012, 181–182) identifies "some form of prohibition of violence, to persons or things, and requirements of truthfulness, fair dealing, and respect for promises" as basic requirements of morality. These standards must be met if living in human societies is to be acceptable, he claims.

158 Narveson (1988, 148) identifies two reasons why humans need morality: (1) because they are vulnerable to others, and (2) because they stand to gain from cooperating with each other.

159 See also Gaus (2021, 59–60), Moehler (2018, 14–15), Stemmer (2013, 44–45).

self-defence. Since individuals would not consent to their creation if they did not exist yet, such moral rules are dysfunctional.[160]

The most inclusive moral community is humanity as a whole. This is the level at which most theories of morality are located. At this high level, however, with such a wide set of addressees, there are only a couple of rules which could be justified by means of a social contract, including e.g. the rule not to kill others except for self-defence. As has been suggested by Moehler (2018), such instrumentally justified higher-level rules can be used as a means of conflict resolution if evolved lower-level moralities diverge.[161] Since they are justified to all rational agents, they may be legitimately applied even in societies characterised by deep moral conflict and also across different moral communities.

3.3.2 *Functional Legitimacy as a Contractarian Approach*

There are many social contract theories in the history of political philosophy, notably those of John Locke ([1689] 2005), Jean-Jacques Rousseau ([1762] 2012), and Immanuel Kant ([1795] 2011). The functional conception of legitimacy, however, forms part of a particular tradition of social contract theory which dates back to Thomas Hobbes ([1651] 1996, 70). What is unique about Hobbes's approach is not the notion of the social contract, but that he relies strictly on a cost-benefit approach to justifying political authority and a stable government, without making further normative assumptions, e.g. concerning individuals' rights or autonomy. For Hobbes, political authority is justified exclusively because it is more in people's interest to have it than to remain in the state of nature. That it is in their interest follows from his modest empirical assumptions of resource scarcity, a universal human desire for continuous preference satisfaction or "Felicity" (Hobbes [1651] 1996, 70), and roughly equal human vulnerability translating into roughly equal strength (Hobbes [1651] 1996, 86–87). In the

160 Among dysfunctional elements of morality which prioritise the collective over the individual are honour codes. For an analysis of how honour codes relate to morality, see Handfield and Thrasher (2019).

161 Building upon his model of *homo prudens*, who has an interest in long term cooperation which outweighs the interest in non-cooperation in any specific case, Moehler (2018, 125) formulates what he calls the "weak principle of universalisation" as a higher-order principle for resolving conflicts among lower-level social moralities. In short, it can be stated as "in cases of conflict, each according to her basic needs and above this level according to her relative bargaining power."

state of nature, without a stable government, these circumstances combined lead to a situation of mutual distrust.

As long as there is no "common Power to keep them all in awe" (Hobbes [1651] 1996, 88), people are therefore miserable, living under the constant fear of violent death in the state of nature which is a state of war of all against all.[162] Importantly, the state of war is characterised by a general disposition to violence rather than by concrete acts of fighting. Due to the total absence of security, the state of war precludes investments in technological progress. People's incentives to leave the state of nature and to seek peace are the prospect to get past the constant fear of violent death, the interest in a better life, and the hope to acquire desired goods by means of labour (Hobbes [1651] 1996, 89–90).

With his interest-based argument, Hobbes initiated the *contractarian* tradition within the broader sphere of social contract theory. What contractarians all have in common is that, starting from modest and purely empirical assumptions, they put forward theories of politics and/or morality which address the problem of long-term peaceful cooperation and argue in terms of individual costs and benefits.[163]

A comprehensive contractarian theory close to the spirit of Hobbes has been developed by public choice economist James Buchanan ([1975] 2000).[164] I will discuss his two-stage social contract in more detail in the following chapters. Another economist and game theorist, Kenneth Binmore (1994, 1998), has worked out an evolutionary contractarianism. Distinguishing between the game of life and the game of morals, he aspires to provide both an explanation and a justificatory criterion of social practices. Ryan Muldoon (2016) also puts forward an evolutionary social contract theory based on Hobbesian assumptions. And Jan Narveson (1988) uses Hobbesian contractarianism as a basis to argue for libertarianism. Without using the label "contractarian", Hart ([1961] 2012, 193–98) also employs a

162 As Narveson (1988, 136–137) points out, Hobbes does not assume human beings to be antagonistic, i.e. to aim at harming each other. Their individual aims are merely contingently conflicting. It requires rules for them to coexist in peace, but peace is not an impossibility.

163 I do not count David Gauthier (1986) among the contractarian camp, even though he identifies his own approach as Hobbesian. Whereas his theory aims at the realisation of cooperative benefits, he does not pay sufficient attention to the issue how cooperative social practices can be stable equilibria. Thus, Gauthier does not demonstrate any interest in the crucial issue which is troubling Hobbes, namely securing peace.

164 As G. Vanberg (2018, 636) points out, due to its commitment to unanimity, public choice theory qualifies as "a modern version of contractarianism."

Hobbesian argumentation in identifying the minimal core of natural law based on empirical truisms about human vulnerability and the possibility of cooperation.

John Rawls, however, while also using the thought experiment of the social contract, is not a Hobbesian contractarian. Rawls's account falls in the *contractualist* tradition of social contract theory. Whereas contractarianism relies exclusively on individual interests as a basis of justification, contractualism allows for normative premises. Accordingly, the "original position" from which he derives his principles of justice serves to illustrate certain restrictions which Rawls (1971, 138) believes should obtain in choosing principles to guide the design of formal institutions. This is achieved by means of the *veil of ignorance* which obscures to individuals their personal identity and preferences (see also 4.4.2). Moreover, Rawls (1971, 19–20) also grants an influential role to normative intuitions by employing the method of the *reflective equilibrium*. This tool requires that in identifying the principles of justice, one iteratively adapts both one's intuitive convictions and the design of the original position. Both the veil of ignorance and the reflective equilibrium are clear indicators that Rawls belongs to the contractualist rather than the contractarian tradition.

Apart from his contractualist starting point, moreover, it is to be noted that Rawls does not even address the question whether political regimes can be justified in terms of benefits. Rawls (1971, 4) already starts out with an understanding of society as "a cooperative venture for mutual advantage" in which the division of labour creates a net benefit of material gains. Presupposing that human beings benefit from cooperating in society, Rawls develops a theory of how formal institutions ought to be designed such that cooperative benefits are distributed justly. This is a very different issue than the problem of political authority tackled by Hobbes (see also Kavka 1986, 182). As Moehler (2024, 28) aptly observes, "Hobbes is not in the justice business, but in the peace business, which aims to maintain a mutually beneficial social order."

The question how individuals can overcome strategic obstacles and cooperate with each other, which is at the core of contractarianism, barely plays a role in the contractualism of Rawls and his disciples.[165] The two traditions thus talk mainly across purposes, not least because they address different problems. Whereas Hobbesians consider peace as the *conditio sine*

165 For an attempted synthesis of public reason contractualism and public choice contractarianism, see Vallier (2018b, 120).

qua non for any other state function such as providing justice,[166] Rawls's take on political legitimacy, in contrast, is that a political system with a nearly just constitution may in consequence have legitimate authority (Rawls 1971, 363).[167]

3.4 The Merits of Hypothetical Consent

3.4.1 *Fair Play*

The functional conception of legitimacy is what Simmons (1993, 76) would call a "quality of government theory,"[168] drawing not on people's actions such as giving actual consent but on the merits of the regime (or other institution) in question.[169] Functionality is characterised by individuals realising mutual benefits, which can be illustrated by a unanimous hypothetical contract. Actual consent, in contrast, has the function of granting rights and incurring commitments. The functional conception of legitimacy presupposes that those people who accept their roles as citizens (or permanent residents) already have certain rights and commitments by virtue of their playing by the rules of the game. Hypothetical consent is not a means of entering the game; it ensures that the rules of the game are fair.

166 Hobbes ([1651] 1996, 100–101) even explicitly makes the point that the question of justice only arises when a reliable order is established:

> Therefore before the names of Just, and Unjust can have place, there must be some coërcive Power, to compell men equally to the performance of their Covenants, by the terrour of some punishment, greater than the benefit they expect by the breach of their Covenant […]: and such power there is none before the erection of a Common-wealth.

167 Being nearly just in Rawlsian terms is a highly demanding requirement. Andrew Fiala (2013, 189–190), for instance, advocates anarchism based on the observation that actual states do not live up to the ideal of Rawlsian justice.

168 Stemmer (2013, 12) similarly distinguishes between "Handlungslegitimität" and "Seins-Legitimität," i.e. legitimacy *qua* act and legitimacy *qua* being. The former arises from acts of authorisation or consent, the latter from the inherent qualities of a norm (or law). A hypothetical social contract models legitimacy *qua* being.

169 It ought, however, be distinguished from Raz's (1990, 129–31) *service conception* of political legitimacy which is concerned with the bindingness of political obligations rather than the justification of existing institutional requirements. On Raz's account, the normal and principal way to justify an agent's authority is that submitting to this authority enables the subjects to better act in accordance with reasons they have than if they were to pursue these reasons on their own. Raz refers to this claim as the *normal justification thesis*.

In this respect, functional legitimacy connects closely to accounts of (political) legitimacy which are based on the notion of *fair play*,[170] in contrast to actual consent. These accounts invoke a principle of fair play to argue that social practices and institutions creating mutual benefits give rise to obligations to participate in such practices for all those individuals who benefit. This is irrespective of the fact whether individuals asked for these benefits. In other words, the principle of fair play entails that there is an obligation to contribute to public goods and common-pool resources.

Public goods and common-pool resources are both *non-excludable*, i.e. people can have the benefits of consuming them without the need to contribute. The difference between them is that common-pool resources are *rivalrous* in that consumption is limited because it depletes the good, whereas public goods are not. Examples for common-pool resources are the classical commons, i.e. jointly used pastures, but also clean air or fish stocks. These are typical cases in which the *tragedy of the commons*, a cooperation problem, arises (see 2.4.2).

Classical examples for public goods are national defence or lighthouses. Public goods may not be rivalrous, but they nevertheless pose strategic issues of the same kind as common-pool resources. The issue is not that public goods would be depleted but that it is difficult to provide them in the first place, relying merely on private individual action. This is because for every potential consumer, it is individually rational, i.e. a dominant strategy, to take the benefit without contributing. Thus, the provision of public goods gives rise to a cooperation problem.

Whereas common-pool resources require that users restrain themselves for reasons of sustainability, public goods require them to contribute their share to them. In both cases, non-excludability has the effect that individuals lack incentives to cooperate; cooperation is a dominated strategy. Proponents of fair-play accounts claim that obligations to contribute to public goods and common-pool resources can be justified to individuals insofar as they benefit from the existence of the good, irrespective of their actual consent.

Consider for instance the fair play account developed by George Klosko (1987).[171] He argues that there are political obligations to contribute to the

170 Not to be confused with Rawls's (1971, 11–12) *justice as fairness* which is the name he gives to his theory of justice. Fairness in Rawls's context refers to the idea that the principles of justice are chosen under fair conditions.

171 Hart ([1955] 2006) also formulates a fair play account of political obligation.

provision of public goods if two conditions are met: (1) The goods provided must be worth more than the costs they impose on the individual and (2) they must be "presumptively beneficial," i.e. goods that everyone can make use of. Klosko claims that if enough others comply with a set of rules to supply presumptive goods, an individual in this society has an obligation to comply as well. He also argues that there are obligations to comply with rules providing non-presumptive ("discretionary") goods insofar as these are added to a scheme of provision of presumptive goods: If the overall benefits do not exceed the overall costs, Klosko claims, the individual is still obligated to comply with the scheme. On his account, one might argue for instance that a government that provides an infrastructure which benefits only some citizens still ought to be obeyed because it also provides peace, which tremendously benefits everyone.[172]

In a similar fashion (yet without using the terminology of fair play), Ronald Dworkin (1990) argues that there are *associative obligations*,[173] emerging not from contractual agreements or voluntary choice but from social practice.[174] Associative obligations, Dworkin claims, exist within families, among friends, but also between citizens in the state if civil society meets certain standards.[175] Whereas associative obligations are not deliberatively chosen, they require reciprocity. The theory of associative

172 Schmelzle (2016, 171) criticises that Klosko's approach to justify political authority by reference to security is paternalistic because security is an optional end of individuals: not everyone necessarily aims at being secure, and it can thus not be presupposed. As I will argue in 4.2.1, however, peace is fundamental for almost anything people may aim for. For this reason, it can be assumed to be an end for all those who have any ends at all.

173 Horton (1992) also understands political obligations as a form of associative obligations.

174 For a combination of a natural duties conception of political legitimacy with an associative element, see Schmelzle (2015, 120–21). On his account, insofar as natural duties are not directly operative, political institutions are required for the political process to determine a reasonable interpretation. This interpretation of the natural duty is binding for the members of the political order in question *qua* members.

175 Dworkin (1990, 222–30) himself identifies two serious objections against ascribing associative obligations at the level of the state: For one thing, states comprise large anonymous societies which differ significantly from small communities where members show equal concern for each other. Moreover, thinking of the state in terms of community sounds suspiciously similar to nationalist and racist claims. Dworkin attempts to evade these objections by claiming that (1) it is sufficient if the practices of a society reflect what can be interpreted as equal concern and (2) that the best interpretation of political practice is not nationalist. Yet these replies presuppose Dworkin's idiosyncratic notion of interpretation and need not appeal to anyone who does not share it.

obligations thus bears a certain resemblance to accounts of fair play, insofar as voluntariness is not required and benefits from association are mutual and cooperative gains.

I do not claim that anyone has duties merely due to the principle of fair play. Yet I agree with accounts of fair play and associative obligations that institutions must create mutual benefits for all their participants in order for the obligations arising from them to be justified. The important difference between my functional conception of legitimacy and the principle of fair play or associative obligations is that functionality is not supposed to be what creates binding obligations but presupposes them. Obligations are an institutional phenomenon (see 2.5). Their existence is independent from their moral justification. Functionality only implies that existing institutional burdens are legitimate. Both fair play and actual consent theorists, however, consider their respective criterion as grounding, not only legitimising, political and other obligations.

The distinction between creating and justifying institutions is important because it shields the functional conception of legitimacy against charges that, by forgoing voluntariness, it allows for putting people under obligations from institutions or social practices they do not even participate in. A popular allegation against fair play accounts of political legitimacy is that the receipt of benefits is insufficient to justify any obligations to contribute to the provision of public goods (see for example Larmore 2020, 115–18). The claim is that incurring (justified) obligations requires consent. Authors who consider consent as a necessary condition for political legitimacy are known as *consent theorists* in the tradition of Locke ([1689] 2005, 330–331).

A well-known argument for consent was made by Robert Nozick (1974) in his *Anarchy State Utopia*. He claims that providing people with benefits is no equivalent substitute for obtaining their consent. Nozick (1974, 93–95) gives the example of a public address system in a neighbourhood of 365 people. Like a radio station, but locally restricted to the neighbourhood, the system provides news, music, and entertainment. Each day a year, another neighbour operates the system and provides a programme for the other neighbours. Nozick makes the point that the mere fact that all other 364 neighbours accept to operate the system on one day of the year does not oblige any member of the neighbourhood to participate in it. This is independent of how much he or she benefits from listening to the programme played by the other neighbours. Even if doing one's share is worth the benefits for a person, Nozick argues, it is not possible to create obligations by setting up a cooperative scheme which happens to benefit people, with-

out being asked for. In a nutshell, Nozick (1974, 95) claims, "[o]ne cannot, whatever one's purposes, just act so as to give people benefits and then demand (or seize) payment."

Nozick's example alludes to the intuition that consent creates an institutional relationship which makes the rules inherent to the institution binding. Indeed, this is sometimes the case, e.g. between the buyer and the seller of a good or service, or among spouses when they enter marriage. Only after both parties have given their consent does the buyer need to pay the price and the seller hand out the good. And only after both have consented to being married are spouses legally required to care for each other. Setting up a public address system would also amount to creating a new institution and the obligations it entails in the first place. This does not merely take place by benefitting people against their will. Insofar as people do not have any obligations, the question whether their obligations are justified becomes obsolete.

The relationship between a citizen and her government, in contrast, exists prior to and independent from either party's consent. The regime is there already, and its legal order is already binding for most citizens, with deliberate consent only accounting for a minority of memberships. This binding legal order may or may not be legitimate in terms of fair play, but the obligations exist in either case. It is therefore crucial to distinguish between the *existence* of an institution, and the obligations it entails, and its *justification*.

Another important institutional type which entails obligations without consent is the family. It would arguably be absurd to criticise the family for the fact that children do not choose their parents. There is simply no way to make such a choice. Newly born human beings depend on the care they receive by adults, even though they are not in a position to choose their caregivers and consent to being in their custody. This also means that parents have no choice but to care for the children they brought into the world. The fact that the family cannot be consensual does not preclude, of course, that some institutional token of the family may on good grounds be criticised for being patriarchal or abusive. Yet it does not imply that the institutional type is dysfunctional as such. And whether a particular token of the family is functional or not is best determined by mutual benefits rather than by consent.

One may argue, of course, that both in the case of the state and the family, into which we are born, a lack of consent is only justified when it comes to minors. Interestingly, however, this argument is not raised with re-

spect to the non-consensual institutions of social morality.[176] People do not consider their moral obligations less binding because they never consented, and they would also not accept a lack of consent as a valid excuse on the part of a person shirking her moral duties. If the institution in question was mutually beneficial, evading one's non-consensual duty would not be an act of autonomy but merely a violation of fair play.

3.4.2 *Voluntariness*

If the criterion for political legitimacy is actual rather than hypothetical consent, moreover, it seems that no existing regime would count as legitimate. Since only a tiny fraction of the population ever took an oath of allegiance to their regime, consent theorists are committed to philosophical anarchism. This is a strong conclusion which not every proponent of consent may feel comfortable with. A consent theorist who does not want to endorse philosophical anarchism may claim, however, that although consent must be actual, it need not be explicit. Instead, she may also allow for *tacit consent*.

An account of tacit consent is, for instance, provided by John Locke. Apart from *express* consent which requires a unique action, Locke also recognizes tacit consent which may be given merely by owning property within the state's territory, and even by using the state's infrastructure when passing through it.[177] Locke ([1689] 2005, 347–348) considers both tacit and express consent as equally giving rise to the political obligation of obeying the state's laws.[178] He even holds that historically, governments

176 Ironically, the internalisation of social morality's consent requirement for the permissibility of many actions is arguably the reason why many people would call for consent to the regime as an authorisation of the government.

177 Rousseau ([1762] 2012, 246), too, takes the position that as soon as a state is established by means of a social contract, residence amounts to consent to be subjected to the sovereign. In a footnote, however, he makes the important qualification that this only amounts to "free" states; otherwise, individuals may face high costs and sanctions in the case of exit, so that they may be forced to stay within the territory against their will.

178 Simmons (1993, 202–203) therefore diagnoses Locke with conflating consent and fair play theories of political legitimacy in his account of tacit consent. And Pitkin (1965, 999) interprets Locke as endorsing a hypothetical-contract theory where a government's legitimacy derives from its merits rather than from consent. Locke's notion of tacit consent does not qualify as a fair play account of political legitimacy,

have indeed been established by consent (Locke [1689] 2005, 336).[179] This position is only tenable if one considers any participation in a regime, that is compliance with the de facto constitution, as tacit consent.[180]

Locke's notion of tacit consent has faced a fierce rebuttal by David Hume. It would be absurd, Hume ([1748] 1994, 193) writes, to suggest that people tacitly consented to political authority by remaining in their native country if they lack any realistic alternative. He offers the following analogy:

> Can we seriously say, that a poor peasant or artizan has a free choice to leave his country, when he knows no foreign language or manners, and lives from day to day, by the small wages which he acquires? We may as well assert, that a man, by remaining in a vessel, freely consents to the dominion of the master; though he was carried on board while asleep, and must leap into the ocean, and perish, the moment he leaves her.

Notwithstanding Hume's critique, Harry Beran (1987, 28–29) claims that native citizens assume political obligation via tacit consent. This occurs, he maintains, by conforming to the convention that residence amounts to consent to membership once adulthood is reached. Empirically, however, it is dubious whether there exists a convention of tacit consent in any given state.[181] Yet even for cases "such as voluntary immigration, running for public office, and acceptance of high-level public employment" (Kavka 1986, 408),[182] this is far from certain. Although it is clear that people, by

---

however. It is rather a descriptive account of participation from which it cannot be inferred that people benefit.

179 A contrary position is taken by Hume ([1748] 1994, 192–93) who claims that governments never relied on consent but always on force and that consent counts the least when new governments accede to power.

180 According to Hampton ([1997] 2018, 94), participation in a governing convention is a weak form of consent. Such consent may explain the emergence of a state. It is, however, not sufficient to give an account of its legitimacy, she holds.

181 This is also where consent theorist Green (1990, 253–254) takes a wrong turn. He claims that citizenship, like marriage, is socially defined but acceded to by consent. Yet citizenship is not defined in this way. Indeed, Green (1988, 168–169) himself observes that there is no consensus on what counts as consent to the authority of the state, in contrast to many other forms of consent such as in the cases of marriage or organ donation. This is a good indication, I would argue, that consent does not form part of the institutional status of citizenship.

182 Kavka understands these cases as usually being instances of tacit consent. He classifies voting in elections and continued dwelling in a country as unclear marginal cases. At the same time, Kavka (1986, 408) demands that for both explicit and tacit consent, "[…] individuals must have reasonable alternatives, and there must not be

performing these actions, participate in the regime, there is not a general convention that this would amount to tacit consent to the *legitimacy* of the regime.

There are indeed some institutional contexts where a convention of tacit consent exists. For instance, tacit consent may occur at a board meeting, as Simmons (1981a, 77–79) points out. If the board members keep quiet after a proposal is made even though they had the opportunity to raise objections, they tacitly consent to it. Yet residence in a state, Simmons argues, differs dramatically from such a tacitly approved decision in that citizens may not be aware of a choice situation so that they cannot intentionally consent. Moreover, there is no way to object to membership in the state, at least at an acceptable cost.

Beran (1987, 76) also holds that, in addition to their tacit consent to the state, citizens who vote in "free and effective" elections consent to the authority of the particular government elected and are therefore under the political obligation to obey its law.[183] But, as Green (1988, 172) argues, citizens are subject to the outcome of a vote, whether or not they agree to the state's authority. Thus, they may simply decide that it is the lesser evil to vote. Moreover, as Simmons (1983, 799–800) points out, elections are not framed in such a way that citizens would be aware of consenting to anything.[184] Neither has majority rule itself ever been consented to by anybody. All these arguments speak against the idea that citizens consent tacitly to their government. What citizens really do is simply participating in social practices and institutions, such as the de facto constitution. This must not be considered as a justification, however, if we do not want to end up equating de facto with justified political authority (see 3.2.2). In the attempt of compensating for not justifying enough, consent theorists might easily justify too much.

In contrast to tacit consent, *explicit consent* as a foundation of political authority is espoused by consent theorists such as Green (1988) and Simmons (1981a; 1983; 1993; 2009). They claim that tacit consent is not binding

so much manipulation of information as to deprive them of the chance to evaluate these alternatives rationally."

183 Irritatingly enough, a few pages before, Beran (1987, 70–74) rejects what he calls the "democracy version" of consent theory, claiming that voting in democratic elections is neither necessary nor sufficient to establish political obligations and political authority.

184 See also Simmons (1981a, 93–93; 1993, 224).

because citizens are not aware of consenting.[185] This criticism, however, seems beside the point. For instance, if the government decided that from next year on, residence amounts to tacit consent, or that citizens have to explicitly consent to its authority to retain their rights of citizenship, individuals would be very much aware of the consent they would give. But their action would not amount to a justification.[186] Citizens already comply with the de facto constitution, so they will also consent explicitly to the state's authority if required.[187] It would be an unwelcome conclusion to consent theorists that any regime can become legitimate merely by labelling actions such as voting or even residence as instances of consent-giving, even if the government acts coercively. As Hanna Pitkin (1966, 43) puts it:

> A government that systematically harms its subjects, whether out of misguided good intentions or simply for the selfish gain of the rulers, is to that extent illegitimate—even if the subjects do not know it, even if they "consent" to being abused.

Thus, even actual consent is not sufficient to guarantee functionality. Indeed, even forced marriage is established by means of exchanging wedding vows. This is notwithstanding the fact that, by construction, marriage without consent could arguably never be justified. Consent theorists do indeed acknowledge that consent may be forced. Since consent may be given under the influence of power, Simmons (1981a, 77) does not only demand that it be intentional, but also voluntary.[188] For instance, he holds that oaths of allegiance in naturalisation procedures can only be understood as voluntary consent if immigrants were not forced to leave their countries of origin and could choose among a set of different countries to go to (Simmons

185 Green (1988, 170–73), Simmons (1981a, 83).

186 Wendt (2018a, 26–27) claims that it would be coercive if the government established a convention stating that non-emigration amounts to tacit consent. Yet the same problem arises if it asks for explicit consent.

187 Binmore (1994, 72) also argues that explicit consent must not be mistaken for a justification for a regime because individuals merely cooperate insofar as this is in their best interest, given power structures as they are.

188 See also Kleinig (2009, 14–20) who lists three conditions for valid consent, namely voluntariness, knowledge and intention. There is also a resemblance to Kavka's (1986, 396) criterion that consent must not be coerced, i.e. that the other party must not be responsible for the consenter's difficult situation (in contrast to forced consent, which is valid, Kavka claims, insofar as dire circumstances result from external causes).

1993, 219). The position that consent must be actual and voluntary to create binding obligations is known as *voluntarism*.[189]

Voluntary consent is an important social practice, both in the formal and the informal sphere. For instance, a requirement of voluntary consent is among the established rules for medical interventions, employment, the purchase of goods, marriage, as well as physical intimacy. If voluntary consent is lacking, attempts to perform or establish these practices and institutions will end up in bodily injury, forced labour, theft, forced marriage, and sexual harassment. And even if consent is given but coerced, it loses any justificatory force. This is notwithstanding the fact that coerced consent may still create a—dysfunctional—institutional relationship, such as a forced marriage.

It is certainly debatable what voluntary consent consists in.[190] Its function, however, is simple: voluntary consent serves as a proxy for the functionality of commitments.[191] In everyday life, voluntary consent is simply the best indicator that individuals will *benefit* from an action.[192] Voluntarily consenting to an action signals that, all things considered, one expects one's situation to be more beneficial if the action is performed than otherwise.[193] For instance, when I consent to undergoing surgery, I express the conviction that I will benefit from it in the long run, such that I am willing to take the cost of being cut open.

Simmons (2009, 306–307) gives two reasons why voluntarist consent theory is attractive as a justification of political obligations: (1) It conforms to the principle *volenti non fit iniuria*, which also comes to bear with promises and contracts, and (2) it expresses the conviction that individual freedom and self-government are morally valuable. Both reasons can be

189 Concerning public goods, Simmons (1993, 255) claims that whereas they cannot be voluntarily *rejected*, *receiving* them may be either voluntary or involuntary. Only in the former case, an obligation can be justified according to voluntarist standards.

190 Wendt (2016, 38–45) names two sorts of conditions to identify what he calls "genuine consent." One is a condition of being able to give voluntary consent on the side of the consenting party. The other condition is not to violate the consenter's basic moral rights. This, however, presupposes an account of moral rights. For more discussions of voluntariness, see e.g. the contributions in Miller and Wertheimer (2009).

191 Greene (2016, 92–93) similarly defines voluntary rule such that the government does not only claim to benefit its subjects by the exercise of powers but that this is in fact the case and that subjects are also aware of it.

192 Vanberg (2004, 156) claims that individuals' voluntary consent is the only available measure of efficiency from a subjectivist and normative-individualist point of view.

193 This is also pointed out by Munger and Vanberg (2023).

understood in terms of benefits. On the one hand, voluntary consent is supposed to protect individuals from avoidable costs. On the other hand, the freedom to choose voluntarily enables them to pursue their own interests, which is a source of benefit for them. As Hobbes already knew, "[…] of the voluntary acts of every man, the object is some *Good to himselfe*" (Hobbes [1651] 1996, 93, emphasis in the original).[194] Mill (1859, 184) makes a similar observation when he notes that a person's "voluntary choice is evidence that what he so chooses is desirable, or at the least endurable, to him, and his good is on the whole best provided for by allowing him to take his own means of pursuing it."

Ensuring that an act of consent is voluntary, however, can be challenging under real-life conditions.[195] Whereas it might be feasible for the institution of marriage with many eligible alternative partners and the viable option of remaining unmarried, it is difficult to see how consent to a political regime could be voluntary beyond doubt.

Simmons (1981b, 28–29, 2016, 122–123) and Beran (1987, 31) imagine that consent to a regime would be more voluntary if there was the possibility to remain an outsider to the legal order without political and legal obligations.[196] Yet remaining outside a political community is not a choice easily made, as Kukathas (2003, 139–140), who also accounts for the possibility of outsiders, points out. Outsiders who reject citizenship will not even obtain a passport to travel elsewhere. People may thus not dare to forego the rights accruing to citizens, even if a regime is not justified to them.

The problem is that the choice to be an outsider is made given the existence of a regime which changes the options available to individuals. This is where *hypothetical contract theories* come into play. Abstracting away from empirical conditions, they take consent under counterfactual circumstances as the criterion of justification, which is voluntary in a way that actual

194 Hobbes ([1651] 1996, 93) also holds that people cannot voluntarily give up their right to self-defence, simply because it is never in their interest not to defend themselves.

195 Pettit (2023, 214) claims that if individuals are sufficiently motivated to comply with the legal order by the benefits it yields to them, rather than by sanctions, their compliance can be considered voluntary. This is questionable for two reasons. For one thing, sanctions are motivationally relevant even for law-abiding citizens because they have an assuring effect (see 2.4.3). Moreover, individuals may have incentives to coordinate even on a dysfunctional regime, not because they receive net benefits but because they incur lower net costs than they currently do (see 3.2.2).

196 Beran (1987, 103–104) also suggests the creation of "dissenters' territories," where those who deny consent to their native states are free to go to.

consent under real-life conditions of institutional power structures cannot be.[197]

Unfortunately, the idea of hypothetical consent has created a good deal of confusion. A frequently voiced worry is that, in contrast to actual consent, hypothetical consent cannot create binding duties or obligations.[198] Moreover, from the perspective of classical liberalism and libertarianism, contractarianism undermines individuals' free choice to assume obligations by consent.[199] The misunderstanding underlying these charges is that a hypothetical contract does not pretend to *create* any obligations, but only endeavours to *evaluate* the legitimacy of institutional arrangements and the burdens they imply.

The argument for hypothetical contractarianism is thus not that hypothetical consent is a substitute for actual consent as a mechanism of institutional authorisation. Rather, a hypothetical social contract is an evaluative tool for social practices and institutions independent of their historical origin.[200] It was never intended to be even "a pale form of an actual contract" (Dworkin 1973, 501). Instead, it is a thought experiment capturing what would be required of institutions such that individuals could voluntarily consent to them.

3.4.3 *Action-Guidingness*

As I argued in the two preceding sections, on the functional account, institutional legitimacy is captured by the notion of hypothetical consent. Actual consent is therefore neither necessary nor sufficient for legitimacy.[201]

197 Thomas Nagel (1995, 36) even frames unanimous hypothetical acceptability as a substitute for voluntariness which is not attainable in the case of subjection to political authority. In fact, however, the notion of hypothetical consent is a detour. Hypothetical contract theory does not care about voluntary consent per se. They merely use it as a proxy for net benefits.

198 See for example Green (1988, 161–162), Simmons (2009, 311), Waldron (1987, 138–139), Wendt (2018a, 30).

199 See for example Holcombe (2018, 97–98) and Levy (2018, 28). For Holcombe (2011), the narrative of the hypothetical social contract empowering governments by unanimous consent is not more than a cynical euphemism serving the propagandistic tool of ascribing legitimacy to the government.

200 See also Buchanan and Tullock ([1962] 1999, 319), Thrasher (2018b, 215).

201 Dworkin ([1988] 2008, 89) also takes this position, arguing that citizens might both consent to authoritarian regimes and deny their consent to governments which actually deserve it (and would therefore obtain citizens' hypothetical consent).

It is not necessary because existing institutions may be fair in the sense that all addressees benefit. Actual consent is not sufficient because consent may be involuntary, tracking only the participation constraint but not functionality. In this section, additionally, I want to argue that hypothetical consent is also superior to actual consent with respect to informing practical decisions, such as which institutions are worthwhile to keep, which ones should be abolished, and also what direction institutional reform should take.

As no existing regime, possibly excluding the Vatican, can claim the voluntary consent of a substantial number of its citizens, actual consent theorists must be anarchists a posteriori if they want to be coherent. This is indeed the conclusion which Green (1988) and Simmons (1981a) draw from the fact that all existing governments lack actual and voluntary consent. That does not, however, commit them to any political position. The lack of consent itself has no practical implications as to whether a regime should be abolished, reformed, or maintained in its current form. Being philosophical rather than political anarchists, neither Simmons nor Green call for the abolition of all political structures. Yet to make the point that the continued existence of some regimes is acceptable, they need to invoke another criterion than consent.

Simmons (1999, 745–48), for instance, grants that some regimes may be justified to exist on the basis of criteria such as providing basic justice, having a lawful regime or being recognised by their citizens and/or the international community. He insists, however, that political authority can only be justified by virtue of consent of the governed. The functional conception of legitimacy, in contrast, takes the converse view to legitimacy. A functional regime may be very imperfect with respect to the functionality of many of its primary laws and lower-level institutions. Its single essential quality is merely that it provides the means for each of its subjects to lead a better life than they could lead in the state of nature. This basic demand helps to distinguish legitimate from illegitimate regimes without invoking other normative standards.

The functional account thus endorses a minimalist conception of legitimacy (see 4.4.3). This makes it well-suited for demarcating among institutional tokens and types which are functional, and those which are not. Functional legitimacy does not commit us to say that all regimes are illegitimate, given the very basic demand that there are regimes which provide all their subjects with net institutional benefits. The latter is arguably the case at least for liberal democracies. On the other hand, it allows us to take a

strong position on dysfunctional institutions which do not even meet this minimal criterion, calling for concrete changes.

Considering how to go about an institutional token, we should first determine whether this token belongs to a functional or a dysfunctional institutional type. Dysfunctional institutional types have functions such that no token would ever be created by a unanimous social contract. Insofar as no token of such an institutional type can be legitimate, we should raise awareness for the illegitimacy of the institution and demand the abolition of this token, as well as of all other tokens of that type. Non-violent practical measures are also commendable. Think for example of the boycott of products from slave labour such as sugar in the late 1800s, or of the South African apartheid regime.

If an institutional token is an instantiation of a functional type, we need to investigate further whether the token itself is functional or not. If it is dysfunctional, e.g. a token of marriage where marital rape is not a crime, we should advocate the reform of the institution such that one day, it becomes functional. The immediate abolition would arguably lead to disruption and deprive all parties of the chance to reap coordinative and cooperative benefits. This is why contractarians often have been sceptical of reforms and taken a more conservative stance on institutional change.[202] Reforms, however, need not be disruptive. They may also take the form of piecemeal social engineering. This is a cautious and negative approach which aims to correct manifest social problems and to eliminate grievances, rather than pursuing a pre-defined vision of society (Popper [1945] 2013, 148–149).[203]

As the example of marriage in Germany shows, an institution may undergo substantial but gradual changes. Whereas it was in 1969 that adultery was abrogated as a criminal offense, the husband's authority to decide about the wife's occupation and the family's place of residence persisted until the 1970s.[204] In 1976, no-fault divorce was made the standard. Marital rape only became a criminal offense in Germany in 1997. And it took another two decades, until 2017, for marriage to be extended to same-sex couples.

202 Contractarians and like-minded theorists tend to argue that the institutional status quo must be taken as the starting point of reforms if these are supposed to be peaceful and mutually beneficial. See for example Binmore (1998, 348), Buchanan ([1975] 2000, 109), Gaus (2011, 460), Moehler (2018, 162), Munger (2018, 59), Vanberg (2004, 158–160).

203 For the advantages of gradual reform, see also Berman and Fox (2023).

204 This was actually in conflict with the constitution, the *Grundgesetz* (GG), article 3, which determined already in 1949 that men and women are equal before the law.

Gradual changes take time, but they may be profound.[205] Gradualism is thus a conservative approach to institutional change, but it is not inimical to change per se.[206] Rather, it is characterised by a certain attitude to *how* change ought to take place, preferring small, slow and continuous steps.

Gradual changes in formal institutions may take place in interaction with the evolutionary development of informal institutions. For instance, in the wake of profound changes in the social perception of gender roles and partnerships, the breadwinner model of marriage went more and more out of fashion in Germany in the second half of the 20th century, while divorce became progressively more accepted. The reform of German alimony law which was adopted in 2007 only became feasible against this background of erosion in the informal norms forming part of the complex institution of marriage. The reform reduced the amount of alimony to be expected in the case of divorce, making it less attractive for wives to withdraw from the labour market upon marriage.[207]

Evolutionary forces, however, may also be employed strategically by activist groups in the deliberate pursuit of their respective agendas, e.g. the suffragettes campaigning for women's' right to vote or the gay rights movement fighting for the introduction of same-sex marriage.[208] Activists may raise awareness for the dysfunctionality of a social practice, and they may also deliberately undermine particular laws by means of civil disobedience.[209]

Even if an institutional token is functional already, we need not stop there. A functional institutional type, such as marriage in Germany after 1997 (i.e. with marital rape being criminalised) may still include dysfunc-

205 As Chirot (2020, 5–6) points out, the long-term effects of piecemeal reform may be as forceful as revolutions.

206 Oakeshott (1991, 431) holds that from a conservative perspective, changes in formal rules must follow changes in beliefs and social practices rather than vice versa. This is what happened in the case of marriage in Germany.

207 Apparently, however, the reform failed to show the intended effect of incentivising married women's participation in the labour market. For empirical evidence, see Bredtmann and Vonnahme (2017).

208 Kitcher (2014, 145–53) describes how outstanding activists contributed to changing norms concerning the social role of women in the West. Kitcher (2014, 162–65) also discusses the process of homosexuality becoming normalised in social morality and law.

209 O'Connor (2019, 202–5) notes that moral education, even if it does not immediately change discriminatory social practices, may have an erosive effect by changing individuals' other-regarding preferences, making illegitimate institutions more susceptible to being overthrown.

tional subordinate institutions or social practices. Even after all the reforms, for instance, marriage in Germany still shows traces of patriarchy, notably in the taxation of married couples. A practice such as income splitting, in contrast to individual taxation, creates incentives for women to work less (Bach et al. 2011), which makes them more dependent upon their husbands. The very function of income splitting is arguably to support marriages that are organised after the breadwinner model. This is not a function which all actual and potential spouses would accept in a counterfactual choice situation. Dysfunctional subordinate institutions such as these should be removed when reforming an institution that is already functional on the whole.

Moreover, subordinate institutions and social practices may also be dysfunctional tokens of functional types. For instance, it may be functional in principle that married couples are required to live in the same place (at least for their first residence), the function being to restrict the benefits of marriage to couples who actually share a household and their personal lives, ruling out sham marriages.[210] Granting husbands the exclusive right to determine the place of residence, however, is not a functional token of this requirement. To become functional, it may be reformed such that both spouses together must agree on one place of residence. So even for institutions which are legitimate, i.e. justified to exist, there is much room for improvement on the functional account of legitimacy.

Deriving recommendations for improving institutions from the principle of actual consent is much more difficult. Simmons (1999, 770) actually holds that while equally lacking legitimacy on his terms, existing regimes may differ in being "more or less fully illegitimate". A criterion for ranking regimes, however, must be different from consent[211] because consent is binary.[212] Functionality is binary, too, so it does not allow for a ranking either. By differentiating between the levels of types, tokens, and subordi-

210 Whether this function is justifiable is of course debatable.

211 Ironically, this criterion seems to be costs and benefits. Elsewhere, Simmons (1981a, 198–199) claims that it is possible to distinguish between better and worse governments insofar as governments do, with varying degree of success, provide benefits by wielding power and coordinating behaviour.

212 Larmore (2020, 118–19) attempts to formulate an alternative conception: Whereas he ascribes full legitimating force only to express consent, he also holds that legitimacy comes in degrees. He gives the example that states differ in the proportion of their population which give express consent. Yet with respect to a subjected individual, consent remains a binary criterion. The same criticism applies in a weaker form to the conception of legitimacy put forward by Greene (2016) who measures the

nate institutions and social practices, however, it can offer a differentiated response to the question how to deal with particular institutions.

These recommendations refer to the very structure of institutions, not to their mere form. This is a remarkable contrast to consent theories of (political) legitimacy. Simmons (1993, 268), for instance, suggests increasing a regime's legitimacy by means of introducing more voluntariness. For one thing, he endorses political activism with the aim of turning existing states into voluntary political societies by offering the possibilities to consent. He also suggests expanding the options open to citizens, for example by offering different levels of citizenship.

The problem with these suggestions is that they do nothing to improve the regime itself which, under given empirical circumstances, might still be the best option to choose for most people. Most importantly, Simmons does not at all suggest any constitutional provisions for how political authority may be exercised. Yet from a functional perspective, constitutional provisions for the exercise of authority are exactly what distinguishes legitimate from illegitimate regimes. They are also the crucial point where legitimate regimes differ from each other. In the remaining chapters, which focus on the legitimacy of political regimes, I will therefore be concerned with matters of constitutional design.

3.5 Summary

In this chapter, I addressed the question what makes institutions legitimate, where legitimacy is understood to mean that an institution is justified to exist. I introduced a functional conception of legitimacy which takes as its starting point that institutions exist to create cooperative and/or coordinative benefits for their participants. Even though institutions all serve such a function, they do not necessarily create benefits for all their (potential) participants. Insofar as those who do not receive any benefits still incur burdens from an institution's existence, they may make the point that an institution is not justified *to them*. Taking a normatively individualistic position, I formulated a principle of legitimacy according to which an institution is legitimate if and only if there is no individual who suffers net costs from its existence. In other words, everyone who incurs institutional

degree of political legitimacy both in terms of the number of citizens who give actual consent and the government's assessed quality.

burdens must at least be compensated by means of coordinative and/or cooperative benefits.

Importantly, people do not automatically signal that an institution is justified to them if they choose to participate in them. This is because, even though they incur uncompensated institutional costs, the very existence of the institution may have made the alternative of not complying even less attractive. On the other hand, people who do not participate in an institution may still benefit from its existence. Even though they do not acknowledge any institutional duties or obligations, these people may legitimately be sanctioned for failing to participate.

Whether an institution is functional or not cannot be precisely measured because individual costs and benefits are subjective values that are inaccessible from an outside perspective. To get a grasp of an institution's legitimacy, however, we can make use of the thought experiment of a social contract. If there is no reason to assume that any individual who incurs institutional burdens would veto the acceptance of a social contract introducing the institutional token in question in a counterfactual situation, known for political regimes as the state of nature, it can be considered functional. Insofar as the state of nature is imagined without any normative presuppositions, the functional conception of legitimacy can be located in the contractarian branch of social contract theory which broadly follows the tradition of Hobbes.

The legitimacy criterion of functional legitimacy is thus consent, but hypothetical rather than actual consent. If an existing institution benefits each of its participants all in all, consent is not necessary to justify it—even though consent may be required to create a new institutional token of a certain type. Actual consent, moreover, may not even be sufficient to capture the requirement of functionality that all participants of an institution realise nonnegative benefits from it. This is not only the case with tacit consent, but also with explicit consent which is given under existing power structures and institutional circumstances, and therefore not necessarily voluntary.

Finally, actual consent fails to be action-guiding with respect to the question whether a particular institutional token should be abolished or reformed. The criterion of functionality, in contrast, which is measured by hypothetical consent, has clear practical implications. Tokens of dysfunctional institutional types such as slavery are beyond repair and should be abolished. Dysfunctional tokens of functional types such as marriage should be reformed. And functional institutional tokens may be improved

by overcoming residual dysfunctionalities at the level of subordinate institutions and social practices.