

Convergent Ethics
and the
Ethics of Controversy
(CEEC)

1. Executive summary

New and emerging life science and plant science technologies (in the following collectively conceived of as new biotechnologies, abbreviated NBTs) are often controversial among members of the public, or are feared to become controversial, in spite of scientific consensus that they can be used in safe and beneficial ways. An influential view holds, first, that while NBTs may be both safe and beneficial, they are nonetheless rife with ethical problems and this is reflected in public skepticism. Second, it holds that the regulatory framework for NBTs should to a considerable degree reflect this public skepticism. Public skepticism, real and anticipated, has thereby come to play a major role in regulation, funding decisions and developments of research strategies for life science and plant science.

The overall goal of the project is to rethink the ethics and democratic regulation of NBTs. First, the dominant theories of what makes actions right or wrong endorsed by professional philosophers (Kantianism, consequentialism, contractualism, virtue ethics) do not uniformly support the public skepticism against NBTs. Rather, we will argue, they support a convergence framework for the acceptability of NBTs stressing consent, sufficient benefit, an acceptable level of risk, and a fair distribution of benefits and burdens of the application of the NBT in question. Second, empirical work in cognitive psychology has documented important mechanisms that are likely to drive the public skepticism. Facts, including facts about technologies, can come to have relatively settled cultural and political meanings, and attitudes toward technologies, including beliefs about their risks and benefits, can hence act as a marker for membership in specific political-cultural identity groups. Individuals are strongly inclined to use their cognitive abilities to support factual beliefs that support their identities, and to defend against information that is perceived as a threat. This leads to biased information processing, causing individuals to reject facts and evidence, and to accepting inflated assessments of risk. This is known as identity-protective cognition, among various different labels. Third, mainstream theories of democratic legitimacy do not straightforwardly support the view that regulations of and framework conditions for research and innovation should reflect public skepticism towards new biotechnologies when that skepticism is grounded in mistaken factual beliefs or in controversial fundamental values.

In the project we will first develop the foundation and interpretation of a convergence framework for the ethical acceptability of NBTs based on the most widely accepted theories in philosophical ethics. Second, we will contribute new empirical research on identity-protective reasoning in a Danish and European context, elucidating i) how cultural and ideological factors influence beliefs and opinions concerning different technologies, ii) the conditions under which novel technologies can trigger identity-protective cognition, iii) deliberative strategies aimed at decreasing the influence of identity-protective cognition regarding biotechnologies, iv) strategies for science communication aimed at decreasing the

influence of identity-protective cognition regarding biotechnologies. Third, we will address questions that arise pertaining to the democratic governance of research, development and application of NBT. Given that NBTs are controversial because of identity-protective cognition, i) how should we understand public participation? Can the pitfalls of identity-protective reasoning be avoided? ii) How should democracy answer to views and opinions that are central to the identity of citizens, yet the result of biased information processing? iii) How can forums for dissemination of scientific knowledge and for conducting public debate overcome identity-protective cognition and its undesirable polarizing effects? Fourth, the project will test and corroborate the convergence framework by applying it to a suite of NBTs in collaboration with partners from the biosciences. Finally, we will execute a comprehensive plan for disseminating the project's findings to the Danish public.

2. Convergent Ethics and Ethics of Controversy

The overarching aim of the project is to approach the ethical and political challenges of novel biotechnologies in a fundamentally new way that draws upon - and expands - work in ethical theory, cognitive psychology and political philosophy, and involves a close collaboration with a range of NBT research communities.

The scientific community has often been bewildered by the public skepticism or even hostility that sometimes arises towards certain NBTs, and which in some cases seems to persist even when the technologies in question are widely regarded as safe and beneficial by the scientific community. Members of the scientific community have questioned the cogency and legitimacy of such skeptical views, perceiving them as irrational or ill informed. The politically dominant reaction has been that while NBTs might well be both safe and beneficial, they nonetheless raise many ethical worries, and these worries are reflected in the public skepticism towards such technologies. Moreover, the regulatory framework for life science and plant science involving NBTs should, to a considerable degree, reflect the public skepticism, ensuring that research and innovation be kept within the limits that a skeptical public finds ethically acceptable. For example, public engagement and participation in decision-making concerning NBTs is widely held to be an important goal, and is a key ingredient of the European Union's science policy program Horizon2020. The importance placed on public participation by policy-makers, funding bodies and other actors is in agreement with, and to a large extent inspired by, a set of views that have become dominant among scholars in the fields of science and technology studies, and technology assessment (for an overview of the evolution of this set of views and their influence on policy, see e.g. Boerse & de Cock Buning, 2012; Gregory & Lock, 2008; Lock, 2008; Bauer, 2009; and Rowe & Frewer, 2000). Public skepticism, real and anticipated, has

thereby come to play a major role in regulation, funding decisions and developments of research strategies for life science and plant science.

While potential ethical problems and public skepticism concerning NBTs should obviously be taken seriously, we want to approach the ethics of NBTs, the public concerns about NBTs, and the democratic challenges that public skepticism gives rise to, in a new way:

First, unbeknownst to most people outside academic philosophy, in a majority of the cases, the general philosophical theories of ethics seriously considered by professional philosophers (Kantianism, consequentialism, contractualism, virtue ethics) support a convergence framework for the ethical evaluation of NBTs stressing consent, sufficient benefit, acceptable level of risk, and a fair distribution of benefits and burdens of the application of the NBT in question. As a consequence, these ethical theories do not find any serious ethical problems with the technologies in question (see more in WP1, below).

Second, psychologists have documented that public skepticism towards novel technologies stems, at least in part, from a suite of cognitive mechanisms that bias information processing and cause inflated assessments of risks. Recent empirical research has uncovered the critical role that values and social groups play in opinion formation, including how they affect processing of factual information (Kahan, 2015; Taber & Lodge, 2006; Bolsen, Druckman, & Cook, 2014). Facts, including facts about technologies, can come to have relatively settled cultural and political meanings, and attitudes toward technologies, including beliefs about their risks and benefits, can hence act as a marker for membership in specific political-cultural identity groups (Lewandowsky & Oberauer, 2016; Druckman & Bolsen, 2011; Kahan et al., 2010). The prospect of having to change one's mind about the risks of a technology can thereby be a threat to part of one's social identity (Sherman & Cohen, 2002). Identity-protective cognition is cognition that aims at defending our identities in the face of such threats. It denotes biased processing of information about identity-relevant beliefs that increasing the likelihood of coming to the conclusion that is congruent with one's identity, even when this comes at the cost of accuracy (Dawson, Gilovich, & Regan, 2002). As an example of identity-protective cognition, people will tend to spend more time assessing evidence *against* their desired conclusions, and will use their reasoning to look for flaws; when assessing evidence in *favor* of a prior conclusion, reasoning primarily aims to support that the evidence is good (Taber & Lodge, 2006). Identity-protective cognition can lead to significant inaccuracies and errors in assessing factual evidence, such as believing that statistical evidence supports the opposite conclusion of what it in fact does (Kahan et al., 2017). Thus, people utilize their cognitive capacities not just to ascertain the risk of a technology as accurately as possible, but also to defend the belief valued by the groups with which they identify. This explains why a common psychological makeup can result in polarization among the public even when they are presented with the same body of evidence – expert testimony, media reports, etc. – concerning an issue, and why this polarization is

greatest among those who are most informed and possess the greatest cognitive capacities (Kahan et al., 2012; Taber & Lodge, 2006; Kahan, 2013; Hamilton, 2011).

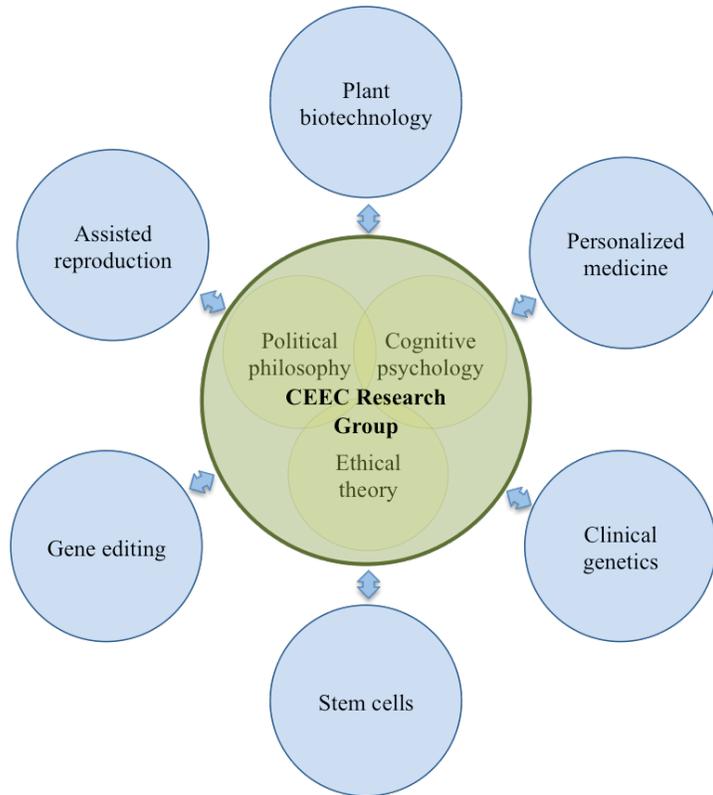
Third, these psychological findings regarding the way that we form normative and factual views about controversial NBTs raise important questions about the nature of democratically legitimate governance of research, development and application of NBTs. As mentioned above, a widely shared view is that democratically legitimate governance should reflect popular skepticism. However, considering the dominant theories in political philosophy detailing principles of democratic legitimacy, it is far from obvious that they support giving considerable weight to political views based on factual assumptions that are at odds with our best evidence, just as it is not clear that normative views that are esoteric in the sense that they are not supported or even rejected by the dominant ethical theories should be given significant democratic weight.

Thus, the overall purpose of the project is to rethink the ethics and democratic legitimacy of new and emerging biotechnologies. More specifically, the objectives are:

- (1) To explore the foundation and interpretation of a convergence framework supported by mainstream ethical theories that we think exists in favor of utilizing NBTs in life science and plant science;
- (2) To contribute to an empirically well-founded understanding of the nature and structure of public skepticism to these technologies;
- (3) To analyze the foundations of democratically legitimate decision making in this context;
- (4) To provide systematic ethical reviews of a range of new and emerging technologies and approaches in life science and plant science done in collaboration with leading experts from life science and plant science;
- (5) To execute a comprehensive dissemination plan targeted at the Danish public in general, policy makers, science communication professionals and scientists.

Each research objective includes a number of sub-questions that are original and theoretically interesting in their own right, but above all the combination of the four objectives is groundbreaking and highly socially relevant. To the best of our knowledge, no similar approach has been developed in bioethics.

The project organization is novel in a number of ways. First, the CEEC research group will collaborate with a handful of NBT science partners from a variety of disciplines (gene editing, personalized medicine, stem cells, plant biotechnology, clinical genetics, and assisted reproduction) rather than focus on just one kind of technology, as is common in similar projects. As the ethical issues and public concerns being raised in these areas are similar to one another, this will provide a much more efficient organization of labor, where more attention can be given to generic issues that are important for a number of different applica-



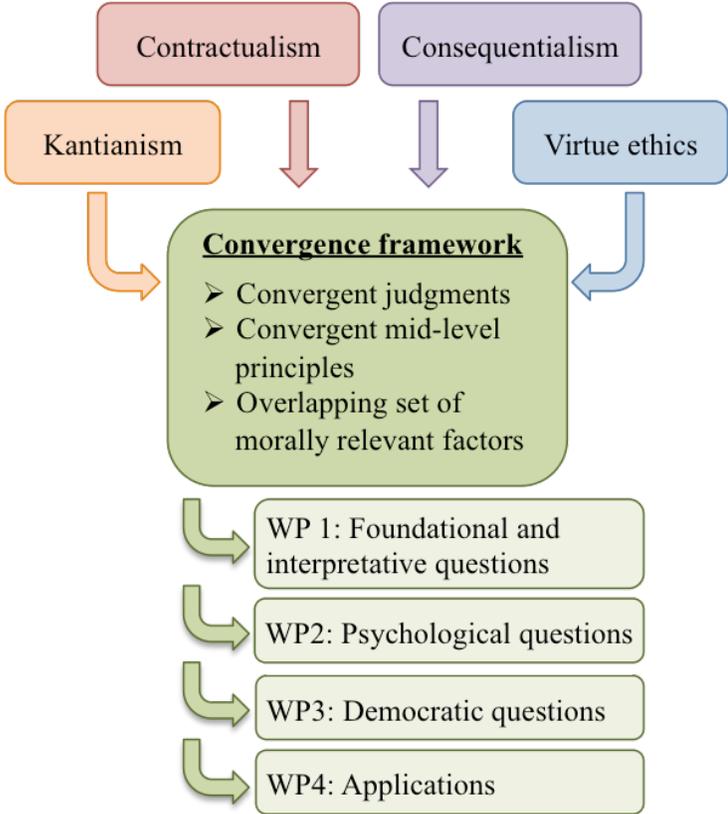
tions of NBTs, and where we can learn from the analysis of similar issues arising across different scientific domains. Moreover, the collaboration with NBT scientists (described in more detail in WP4) will enable us to test and develop the convergence framework in a robust way, and will secure a much more efficient dissemination in the relevant scientific communities. It will be easy for the group to expand activities to others fields in life science or plant science. Second, the project is also novel in that it combines research in ethical theory, cognitive psychology and political philosophy in ways that are obviously relevant, though these disciplines are rarely brought together in one research project focusing on a complex problem such as democratic governance of NBT.

Finally, the project is novel in that it will involve mostly collaborative work. The majority of the planned papers will be co-authored by members of the group and collaborators. Jointly authoring papers brings clear advantages of facilitating critical discussion, sharing of expertise and know-how, and distributing of cognitive labor, but is still rather rare in philosophy. The members of the group are trained in this way of working due to their participation in earlier cross-disciplinary projects (see section 5).

Apart from producing novel and socially important high-quality research aimed at academic audiences, the project includes a comprehensive dissemination plan that targets the Danish public at large; high school students; science communication practitioners; bio-science researchers, and policy makers. The project thereby aims to have a lasting impact

on Danish public debate, policy-making and science communication. Moreover, we hope that the project will contribute to the design of communication strategies and procedures for democratic engagement that better meet the objectives of democratic legitimacy, and that avoid turning the debates on new technologies into mere proxies for a battle between different cultural identity groups.

Overview. The project is divided into four work packages described in Section 3 below, each corresponding to one of the objectives stated above. Work Package 1 develops what we call a convergence framework for NBTs. Work Package 2 contributes new empirical research on identity-protective cognition in a Danish and European context. Work Package 3 addresses a range of questions in political philosophy that arises in so far as identity-protective cognition contributes to citizens holding false or unjustified factual beliefs, and moral beliefs that fall outside the scope of the main theories. In Work Package 4 we apply and test the framework with our col-



laborators from biosciences. Section 4 of this document describes an ambitious dissemination plan targeting the Danish public. An overview of national and international research interaction is provided in Section 5, while the governance of the project is detailed in section 6. A detailed timeline of all deliverables and activities in the project, and an allocation of resources to the various research tasks found in Section 7. In the description of the work packages below we have added projected papers with tentative titles and proposed publication venues. While most papers will be jointly authored, we have assigned a lead author to each of the projected papers to ensure clarity about distribution of tasks.

3. Work Packages

WP1: Convergent Ethics: Developing a convergence framework for assessing novel biotechnology

The aim in WP1 is to provide a general characterization and in-depth analysis of the foundation and the interpretation of what we will call the convergence framework. We expect WP1 to be supporting at least three kinds of interesting general claims:

- (1) There is a convergence among major ethical theories in the evaluation of NBTs, which can be expressed in terms of a set of mid-level principles along the following lines: NBTs are ethically acceptable in so far their use involve
 - (a) the right form of consent by affected individuals,
 - (b) sufficient benefit to users and/or society at large,
 - (c) acceptable risk, including an acceptable distribution of risk across individuals,
 - (d) fairness in distribution of benefits and burdens imposed by the technology.
- (2) The convergence excludes a range of factors that public opinion takes to be ethical concerns par excellence (e.g. concerns over unnaturalness and ‘playing God’, general feelings of unease and repugnance, absolutist ideas (e.g. about the sanctity of life), and extreme or exclusive focus on possible (but unlikely or even far-fetched) bad outcomes).
- (3) As a result, major ethical theories converge in a generally positive assessment of a range of NBTs, including the cases where they are frequently considered ethically controversial by popular opinion.

WP1 is subdivided in two parts: one that addresses foundational questions concerning the nature and epistemic significance of convergence in ethical theory, and one that concerns how best to interpret the content of the mid-level principles and related questions.

1.1 Foundational questions: The nature and epistemic significance of convergence in ethical theory

Our aim in this part of WP1 is to analyze the nature and the rational significance of the theoretical convergence we posit. The main hypothesis is that major ethical theories converge in their assessment of a range of novel biotechnologies broadly understood, and that they are generally less negative (or not negative at all) compared to public opinion about the more controversial cases. Thus, there is a significant divergence between how ethical theories tend to evaluate NBTs and public opinion concerning them. By *ethical theory* we here mean a general, systematically elaborated philosophical views of what we morally ought to

do. By *major* ethical theory we have in mind those theories that have been and still are dominant in the Western philosophical tradition: consequentialism, Kantianism, contractualism and virtue ethics. e.g. Kantianism (e.g. Kamm 2007; Nozick 1974), consequentialism (e.g. Broome 1991; Kagan 1989), contractualism (e.g. Gauthier 1986; Scanlon 1998) and virtue theories (e.g. Foot 1978; Hursthouse 1999) and their later developments. Each comes in many variations, they might be partly overlapping, and there might be views that combine features from the mainstream views.

Following Kagan (1992, 1998) we can distinguish between ethical theories at the *factoral* and the *foundational* level. The factoral level concerns what the set of *morally relevant factors* (MRFs) are (or, equivalently, what considerations have genuine reason-giving weight), as well as their relative weights and more precise explications of the content and limits of factors. Foundational theories provide an analysis of the concept of moral rightness and (thereby) purports to explain what factors are morally relevant and why. We can now distinguish different conceptions of convergence: (a) Two theories may converge in the sense that they assert the same set of MRFs or a large enough intersection of MRFs, at least as far as the evaluation of NBT is concerned. (b) Alternatively, two theories may converge in their evaluation of a range of NBTs in that they imply the same overall moral verdict in a certain range of cases, or agree in their implication regarding moral status, whether or not this agreement in moral evaluation is based in the same set of MRFs. (c) A third option is that two ethical theories converge in the sense that they support the same mid-level principles (or practical guidelines/policy guidelines) regarding the assessment of NBT in a wide range of cases, even if they assert a different set of MRFs. A variant of this is that two theories support the same mid-level principles when they are stated in rather general terms, though they do not necessarily agree on the detailed interpretation of the mid-level principles. In Rawls' terms they agree on certain concepts, but not necessarily on specific conceptions. In some cases, different ethical theories may display near-convergence, which is situation where any divergence of views about ethical acceptability hinges on a single controversial premise (e.g. whether nature is intrinsically valuable). Such near-convergence may be illuminating, since widely held theories of democratic legitimacy place significant weight on whether disagreements on policy are based on controversial metaphysical, ethical or religious views (Gaus 2011; Rawls 1993). We will explore all these kinds of convergence. In particular the third kind (c) is most relevant for policy-makers and scientists, as they may support useful practical guidelines and policy-making tools.

Intuitively, broadly defined technologies or approaches such as DNA sequencing, CRISPR or synthetic biology are not the appropriate objects of ethical evaluation. Rather, it is the specific application of such technologies for a particular purpose and in a specific legal and social setting that we should be concerned about, not the technologies or approaches as such. In part this is a result that we expect to be corroborated in our analysis. The reason is that if very generic objections to NBTs were accepted (such as naturalness,

playing God, taboos...) then much more general evaluations of NBTs would make sense. But since we suggest that the major theories converge on views that leave no room for these, we also think that the evaluation should be focused on specific applications of technologies.

Convergence and near-convergence in ethical theories regarding the evaluation of NBTs are at least significant in the sense that they contradict what is probably a widespread opinion among scientists and policy-makers that NBTs, while potentially beneficial, are nonetheless ethically problematic. This statement, if not outright false, is surely in need of significant qualifications. Second, one would think that it would constitute significant evidence that dominant but yet diverse approaches in ethics converge: ethical experts with a plurality of backgrounds agree on these matters, so it is very likely to be correct. Clearly, however, this question is not straightforward, as is shown in recent discussions of the significance of disagreement and consensus in philosophy (see e.g. Christensen and Lackey, 2013). We want to discuss the most important reasons for saying that convergence is epistemically significant, and the most important reasons for saying that it is not.

Expected outcome

Paper 1.1: How major ethical theories converge on criteria for the ethically acceptable use of novel biotechnologies

Paper 1.2: The rational significance of convergence, near-convergence and disagreement in ethical theory

Lead author: Klemens Kappel. Publication venues: Applied and general ethics journals, e.g. Ethical Theory and Moral Practice, Ethics.

1.2 Convergent ethical principles and their interpretation

The aim of the second part of WP1 is to analyze in more detail some of the questions concerning how the convergent principles may reasonably be interpreted. We will focus on the interpretations of condition (a) regarding consent, condition (b) concerning benefit, and condition (c) concerning risk. In all cases, we will, for the purpose of argument, consider the plausibility of fairly minimal interpretations of these conditions. Finally, we will analyze in depth what may appear to be a problematic type of cases for the convergence framework, namely cases of exploitation.

1.2.1 Consent. One of the convergent principles hypothesized is that morally acceptable application of NBTs requires an appropriate form of consent from affected individuals. Obviously, this leaves open large questions of interpretation (for general literature on consent see e.g. Kleinig, 2004, Nielsen, 2004, Hurd, 1996). Informed consent to a medical in-

tervention or procedure is of critical importance in medical ethics, as a means of protecting and promoting patient autonomy (see e.g. Faden & Beauchamp, 1986, Beauchamp & Childress, 2009). As such, informed consent serves an important moral function in medical practice. The paradigmatic medical case standardly assumed to require informed consent features a competent patient directly involved in a medical procedure (say, by receiving an injection or surgery), where the direct consequences of the intervention affect the patient only. Many applications of NBTs deviate from this paradigm: Often applications of NBTs do not affect users in the same immediate way as injections or surgeries do; the consequences of using various NBTs may be distributed over many individuals (including future individuals); or there may be no particular consequences for individuals. In other cases, non-users of NBTs may be affected in significant ways that differ from what one sees in the paradigmatic medical case.

Consider large state-operated biobanks that are integrated in national health care systems that provide statistical data required for improving medical treatments. One might say that the fact that a biological sample from a particular individual is included in the database does not affect this individual in ways similar to the paradigmatic case (but may affect her in other ways), and the inclusion of the sample in the database standardly has no direct consequences for the donor. Now, consent schemes for biobanks are usually in place out of concern for issues such as secondary and future uses of the sample, privacy, and benefit sharing. Nevertheless, it is clear that such issues are of a different nature than in the paradigmatic medical case in that there are no direct physical consequences for the individual. One might ask what sort of consent the major theories/convergent principles support in such cases. Is a requirement individual specific informed consent supported by the major theories? Is even a requirement of broad consent supported (Lipworth et al., 2006; Eyal, 2012; Tutton, 2001; Kaye et al., 2015)? Maybe what matters in such contexts is not informed consent (narrow or broad), but rather that the decision to participate (or, to use a given NBT) is made upon trust. At least it is evident from a still growing body of empirical studies that patients, donors and research subjects across the globe do not place as much weight on the information they receive about an intervention, procedure or study, as they do on the trust they hold in the research institution, researcher and or doctor, when deciding whether or not to consent (Nobile et al., 2016; Wadmann, 2013; Høyer, 2010; Molyneux et al., 2005; McDonald et al., 2008; Kass et al., 1996). This challenges the moral weight informed consent is accorded in medical ethics and practice, and raises the philosophical question of whether trust may protect and promote the autonomy of patients, donors and research subjects to the same, or perhaps even larger extent as information is held to do. We will explore and develop the idea that consent based on trust is not morally inferior to consent based on information, given certain boundary conditions that we identify (Kongsholm & Kappel, 2017; Manson & O'Neill 2007). Note that this is compatible with holding that

there will often be pragmatic or instrumental reasons to have more demanding regimes of consent.

These questions obviously have important implications for the governance of bio-banks. But how should we think of consent in the context of other NBTs? In some cases it is clearly true that while the application of some technology or intervention does *not* directly involve a particular individual, he or she may nonetheless be significantly affected. Enhancement technologies illustrate this. If some members of society use enhancement technologies to significantly enhance intelligence, this may significantly affect those members of society that do not. Something similar may be true for gene-editing technologies, or even the use of plant biotechnologies: even those who are not directly involved may nonetheless be affected in other ways that cannot generally be dismissed as insignificant. We want to discuss the limiting cases of being affected by the use of a NBT that generate a moral requirement of consent.

Expected outcome

Paper 1.3: Trust, mistrust and consent in novel biotechnologies

Paper 1.4: To inform or not to inform: when is consent based on trust enough?

Paper 1.5: When should you consent to something that does not affect you directly?

Lead author: Nana Kongsholm (postdoc). Publication venues: Bioethics and ethics journals, e.g. Bioethics, Journal of Medical Ethics, The Journal of Ethics.

1.2.2 Benefit. Empirical studies have shown that people generally approve more of medical applications of genetic modification than of applications in food production (Biotechnology and the European Public Concerted Action Group, 1997; Gaskell et al., 1999; Olynk Widmar et al., 2017). An oft-suggested interpretation of this finding is that the potential benefit of a technology should be high for it to be ethically acceptable, and citizens find that the potential benefit is high in the case of medicine, and low in the case of food - e.g. because we (in the affluent West) have enough food, or because ‘natural’ foods (but *not* natural medicines) are believed to be more healthy than non-natural alternatives (Rozin et al., 2004). This raises the question of what benefits NBTs need to deliver if they are to be ethically acceptable. Our assumption is that only a relatively low level of benefit is needed, at least as long as the technology is safe. We further assume that benefits should be measured in a way that is compatible with standard liberal requirements concerning state neutrality about the good life. The main interpretative issue concerns this latter assumption: What exactly does liberal neutrality entail with respect to NBTs? Our hypothesis is that neutrality can only very rarely ground *restrictions* of NBTs, but will rather support a relatively low bar for sufficient benefit.

1.2.3 Risk. Risk is among the most invoked objections to NBTs, and risk management has played a crucial role in arguments for ‘democratizing’ the governance of technology (cf. section (3.3) below). We will explore what the convergent principle of acceptable risk entails. In particular, we will suggest that the notion of adequate precaution plays an important role in this principle (others, e.g. Resnik (2012) and Sandin (2015), have made the similar suggestion that precaution itself is a mid-level principle). In previous work (Christiansen, 2016, ch. 3) we have argued that there should be convergence from core theories (including consequentialism) on the use of threshold conditions of harm and uncertainty in legal/administrative versions of the precautionary principle, and that this convergence is justified both by moral and rationality-based reasons. We will further explore the possibility of unifying the moral and rationality-based ideas that are involved in precaution (as urged by Munthe, 2015). In particular, we will investigate the relation between two core intuitions behind precaution, namely (a) that acting to prevent possible harm can be justified on the basis of limited evidence (as stressed e.g. in the Rio Declaration), and (b) that acting to prevent a possible harm can be justified if the possible harm is disproportionate to the possible benefit, even where the harmful outcome is much less likely than the benefit (Randall, 2011, ch. 8).

Expected outcome

Paper 1.6: How beneficial should NBTs be to be morally acceptable?

Paper 1.7: Rationality and morality in the justification of precaution

Lead authors: Klemens Kappel and Andreas Christiansen (postdoc). Publication venues: Bioethics journals, e.g. Hastings Center Report, Bioethics.

1.2.4 Exploitation. Setting aside details of interpretation of the convergence framework, it might seem uncontroversial that the framework plausibly identifies a set of ethically relevant features for the assessment of NBTs, and screens off others. It is important to note, that there is a set of cases that at least initially seem to challenge the framework: Some interactions between individuals may involve consent, mutual benefit, and acceptable levels of risk, and yet not be ethically unproblematic. These cases are often couched in terms of *exploitation* (Wertheimer, 1996; Feinberg, 1988; Hill, 1994). In the literature, exploitation is particularly discussed in relation to practices such as commercial surrogacy, clinical research, and biobanking. In these practices, there is a risk that surrogate mothers, research subjects, or biobank donors are exploited by in various ways. It is not immediately clear that the convergence framework would capture this.

There are two dominant types of theories of exploitation. One understands exploitation as *unfairness*. On this view, an otherwise mutually consensual interaction is exploitative if it results in a highly unequal or otherwise unfair distribution of the risks and benefits

involved (Wertheimer, 1996; Valdman, 2009; Wilkinson, 2003). From this perspective, a research subject is exploited if she is not sufficiently compensated for her enrollment in a research program, or if the compensation is not proportionate to the risks involved (El Setouhy et al., 2004). The other dominant type of theory links exploitation to *vulnerability* (Goodin 1987; Wood 1995). According to these views, exploitation may result when interacting with vulnerable individuals, e.g. because such individuals may enter relations on a less than fully voluntary basis. Vulnerability may here be understood as an individual being in a difficult or precarious situation, or an individual being impaired in her decision-making capacities. Vulnerable individuals are prone to be used by others in way that may constitute exploitation, even if they have in some sense consented to the interaction. Exploitation in this sense cannot be countered by simply allocating a larger share of benefit to the exploitee, as the first type of theory above seems to imply. Rather, someone may be exploited in a relation regardless of the amount of payment she receives, and sometimes, too high a payment may in fact render a relation exploitative, as it can serve to undermine consent. For example, it is a general concern in research ethics that a high level of compensation for research enrolment in resource-poor settings may be considered to be exploitative as it can constitute an undue inducement of a vulnerable individual (Dickens & Cook 2003; CIOMS 2016).

This part of the project aims to elucidate the ways that various applications of NBTs can be claimed to be exploitative, or imply exploitative practices, in the light of recent philosophical work on exploitation (e.g. Wertheimer, 2011; Valdman, 2009; Sample, 2003; Wood, 2016). Understanding the potential or alleged potential for exploitative practices across a range of different NBTs is interesting in its own right, but it will also function as test case to the plausibility of the convergence framework. Thinking about exploitative cases may lead to modification of the framework, or press in the direction of particular interpretations of it.

Expected outcome

A PhD dissertation including papers with the following tentative titles:

Paper 1.8: (How) can new biotechnologies be exploitative?

Paper 1.9: Exploitative cases: ‘invalid consent’ revisited

Paper 1.10: New biotechnologies: when does unequal distribution of benefit imply exploitation?

Lead author: Katla Heðinsdóttir (PhD-student). Publication venues: Bioethics and general ethics journals, e.g. BMC Medical Ethics, Ethical Theory and Moral Practice, Journal of Applied Philosophy. Klemens Kappel will supervise the dissertation.

WP2: Cognition and deliberation about biotechnologies

WP2 will consist of empirical work, which aims at contributing to our knowledge of the cognitive mechanisms underlying public opinion about biotechnologies. In particular, we will focus on a class of cognitive mechanisms that fall under the header of identity-protective cognition. While research is continually unveiling details about the mechanisms of identity-protective cognition (Kahan et al., 2017; Taber & Lodge, 2006; Dawson et al., 2002), it is as yet relatively under-explored what factors determine whether a given technology becomes culturally meaningful in the first place (Kahan et al., 2010; Druckman & Bolsen, 2011). Likewise, relatively little is known about which communicative and deliberative strategies may effectively counteract the influence of identity-protective cognition (Cook, Lewandowsky, & Ecker, 2017; Cohen et al., 2007; Prior, Sood, & Khanna, 2015). Finally, the vast majority of extant research on identity-protective cognition has been carried out in an American context, which may differ in important ways from a European context.

The aim of WP2 is to help fill some of these gaps in our knowledge. Specifically, we will investigate

- (1) which demographic, cultural and ideological factors influence beliefs and opinions and bias information processing concerning different technologies, in particular vaccines and genetically modified organisms, in a Danish and European context,
- (2) conditions under which novel technologies can come to carry cultural meanings and trigger identity-protective cognition,
- (3) deliberative strategies aimed at decreasing the influence of identity-protective cognition regarding biotechnologies; and thereby the polarization of public opinion on matters of fact,
- (4) communicative strategies aimed at decreasing the influence of identity-protective cognition regarding biotechnologies and thereby the polarization of public opinion on matters of fact.

The methods of WP2 will include experiments embedded within large nationally representative surveys, and a laboratory experiment involving both individuals and deliberating groups.

2.1 Culture and identity-protective cognition about vaccines and GM foods

The main aim is to identify cultural variables that explain variance in a nationally representative sample of Danes' attitudes toward biotechnologies, particularly the MMR vaccine, the HPV vaccine, and genetically modified organisms, and to demonstrate the influ-

ence of identity-protective cognition in processing of factual information pertaining to these technologies.

Background. In the United States, public opinion about numerous technologies is sharply divided along political fault lines. For instance, American liberals tend to judge the HPV vaccine as relatively low risk while conservatives tend to judge it as high risk (Kahan et al., 2010), while the opposite pattern is found for nuclear energy and global warming (Kahan et al., 2012; Peters & Slovic, 1996).

While political affiliation predicts opinion reasonably well, psychometric measures of cultural worldviews explain variance over and above demographic and political factors. In particular, work within the framework of cultural cognition theory has shown that people's position on a two-dimensional measure of culture (Douglas & Wildavsky, 1983) predicts beliefs about numerous technologies over and beyond demographics and political affiliation (Kahan et al., 2007). One dimension, individualism-communitarianism, denotes the extent to which people prefer collective solutions to societal problems over individual and market-driven solutions. The other dimension, egalitarianism-hierarchy, describes the extent to which people prefer firmly stratified social orderings in roles and authority. These two dimensions combined predict a large proportion of the variance in the American public's attitude to a multitude of issues. Citizens may perceive factual matters to clash with aspects of their worldview, motivating identity-protective cognition. The example of global warming illustrates this: If global warming were a serious risk, which requires regulation of industry to mitigate, then this would impugn the competence of societal elites and the ability of the market to solve societal problems. This would be threatening to hierarchical individualists, for whom the competence of societal elites and the problem-solving ability of markets are important tenets. Thus, they tend to believe that anthropogenic global warming poses little risk, or to deny its reality (Heath & Gifford, 2006). When faced with evidence suggesting the reality and severity of global warming, they will engage in identity-protective cognition: Using their powers of reasoning to make the evidence yield the conclusion that favors their cultural worldview (Cook & Lewandowsky, 2016). This helps explain why, contrary to what one might expect, the most scientifically literate, most intelligent, and most educated are *not* those most likely to converge on experts' view of the facts about global warming. Rather, they are the ones who are most likely to display ideological polarization. Among hierarchical individualists, those with greatest cognitive and scientific ability are most likely to deny the reality of climate change (Kahan et al., 2012).

Numerous biotechnologies have been found to show a similar pattern, with opinion sharply divided along cultural fault lines and identity-protective cognition at work during the processing of relevant information, including the HPV vaccine (Kahan et al., 2010) and stem cell-research (Nyhan & Reifler, 2010).

While the cultural cognition model has thus been remarkably effective in explaining patterns of American public opinion and its cognitive underpinnings, it is unclear whether it will prove as successful in explaining patterns of public opinion within European countries (Maleki & Hendriks, 2014; Grendstad, 1999). Existing studies on identity-protective reasoning in Europe have tended to focus more on political party identification than underlying worldview variables (e.g. Petersen et al., 2012, cf. von Boorgstede et al., 2014). Likewise, there is reason to suspect that different technologies may be culturally contested in Europe as opposed to the US. One example of this is GM foods: Americans are generally positively disposed toward GM foods compared to Europeans (Gaskell et al., 1999), and there is no evidence that opinion is divided among political or cultural lines.¹ In contrast, the sizeable European opposition to GM crops is greatest on the political left and smallest on the right (Durant & Legge 2005). Another example is the HPV vaccine, which is culturally contested in the United States, with hierarchical individualists tending to view it as carrying high risks while liberals tend to view it as low risk (Kahan et al., 2010). While the HPV vaccine is indeed highly contested among citizens in Denmark, the debate over the vaccine does not seem to be as overtly political as was the case in the United States, and the cultural elements of the public resistance may be markedly different. We aim to contribute to extant knowledge by investigating the cultural cognition of biotechnologies in a European context.

Hypotheses.

- (1) Cultural variation explains variation in public opinion about MMR vaccines, the HPV vaccine, and GMOs, and does so better than demographic variables.
- (2) Identity-protective cognition biases processing of factual information relevant to the risks and benefits of these technologies.

Methods. We will conduct a large (N≈2000), nationally representative survey through a surveying agency (e.g. YouGov). The primary dependent variables will be measures of ethical attitudes toward and factual beliefs about a range of biotechnologies, provisionally including the MMR vaccine, the HPV vaccine, and GMOs. In order to identify whether public opinion about biotechnologies are culturally divided, we will include cultural measures in addition to standard demographic variables. The most promising scales will be identified in pilot testing, but we expect to include Danish adaptations of the hierarchy-egalitarianism and collectivism-individualism scales (Kahan et al., 2007), as well as a standard measure of political ideology and free market support (Lewandowsky et al., 2013). We will use two methods to identify the presence of identity-protective cognition. First, we

¹<http://www.culturalcognition.net/blog/2013/11/5/we-arent-polarized-on-gm-foods-no-matter-what-the-result-in.html>

will include measures of cognitive ability and scientific literacy (CRT, numeracy scale, Ordinary Science Intelligence (Frederick, 2005; Kahan, 2014b)). If we find that polarization among cultural groups increases along with cognitive ability, this suggests that cognitive ability is being used to defend the view afforded by one's cultural values. Second, and more directly, we will include an evidence-evaluation task in the survey. Subjects will evaluate one or more pieces of evidence (numerical data, an excerpt of a scientific or newspaper article, etc.) relevant to the target technologies. We will measure their evaluation of the evidence, including processing time, and include a comparison of attitudes and beliefs before and after the evaluation of evidence. Identity-protective cognition could be inferred by differential assessments of evidence on the basis of culture (Taber & Lodge, 2006), and on finding that attitudes polarize further after exposure to the same evidence (Lord, Ross, & Lepper, 1979).

Expected outcome

Paper 2.1: Cultural and cognitive underpinnings of vaccine and GMO attitudes

Paper 2.2: Biased assimilation and attitude polarization in response to evidence about vaccines and GM foods

Lead author: Bjørn Hallsson (postdoc). Publication venues: Psychology or political science journals, e.g. Psychological Science, Political Behavior, Political Psychology.

2.2 Culture and novel technologies

The aim is to identify features of technology or external conditions that put novel biotechnologies at risk of becoming culturally contested, and thereby subject to identity-protective cognition among citizens.

Background. While the impact and mechanisms of identity-protective cognition are gradually being unmasked, much less is known about the risk factors and triggers that cause factual questions become embedded in personal and social identities, and thereby subject to identity-protective cognition, in the first place. Finding answers to this question is important, particularly for novel technologies. Identifying the features of technologies or their cultural context that put them at risk of becoming salient to social identities may allow for taking appropriate countermeasures in efforts to protect the factual questions from the distorting effects of identity-protective cognition.

Features of technologies themselves may be a risk factor if they can easily be perceived as relevant to a broader cultural framework. For instance, consider the difference between opinions toward the MMR vaccine and the HPV vaccine in the US. The MMR vaccine has largely escaped political polarization in the US, and opposition to it, while vo-

cal, is very small (Kahan, 2014c). On the contrary, perceptions of the risks of the HPV vaccine are politically and culturally divided. One proposed reason for this difference is that the HPV vaccine, by virtue of protecting against sexually transmitted pathogens, was easily brought to bear on a wider cultural conversation about sexual ethics. Conservative critics were concerned that the vaccine would amount to a promotion of teenage promiscuity (Charo, 2007).

External conditions may also be important determinants of identity protective reasoning. These include how a technology is framed in the wider debate, which risks and benefits are highlighted and which values are promoted, who the advocates of the technology are perceived as being culturally aligned with, as well as the legal and institutional framework surrounding the technology (Druckman & Bolsen, 2011; Kahan et al., 2009; Kahan et al., 2015). In the case of the HPV vaccine, the controversy took broad hold only after public outrage about what was perceived as a rushed process of FDA approval and a subsequent push, apparently influenced by lobbying from the vaccine manufacturer, by some government actors to make vaccination mandatory (Blumenthal, 2007; Elliott, 2007; Kahan et al., 2010).

Technologies can be vulnerable to identity-protective cognition without in fact being widely contested. This will often be the case when the technology is relatively unfamiliar. For example, opinions about nanotechnology were found to be identical for hierarchical individualists and egalitarian communitarians when they were given only a short description of the technology, but to sharply divide after exposure to two paragraphs detailing the possible risks and benefits of the technology. Egalitarian communitarians focused on possible environmental and health risks and largely ignored possible benefits, becoming much more negative toward nanotechnology, while hierarchical individualists credited possible benefits and largely ignored the possible risks, and became more positive (Kahan et al., 2009).

To the best of our knowledge, no study has tested whether a novel technology can be perceived in culturally opposing ways depending on the risks and benefits highlighted in an accompanying frame.

Hypotheses.

- (3) Framing of an unfamiliar technology as culturally salient causes opinion to align along cultural lines and for identity-protective cognition to bias subsequent processing of information about risk.
- (4) The same technology can be aligned with different cultural groups dependent on framing.

Methods. We will embed an experiment in the survey outlined in (2.1). Therefore, all participants will fill out the demographic and cultural scale items described previously. Our dependent measures will consist of ethical attitudes toward and factual beliefs about a novel biotechnology (e.g. CRISPR/Cas9) which participants can be expected to be relatively unfamiliar with. For one third of participants, indication of attitudes and beliefs will be preceded by a culturally neutral description of the technology. For another third, it will be preceded by a description of the risks and benefits of the technology that we expect to align it with communitarian and egalitarian values, and for the final third, it will be preceded by a description of the risks and benefits of the technology that aligns it with hierarchical and individualist values. We then present the attitude and belief measures again. Our hypotheses will be corroborated if we find that information about risks and benefits causes cultural polarization in different directions depending on the cultural affinities of the risks and benefits highlighted in each condition.

Expected outcome

Paper 2.3: Creating a toxic science communication environment: Identity-protective cognition about emerging technology.

Lead author: Bjørn Hallsson (postdoc). Publication venues: Psychology or communication journals, e.g. Journal of Communication, Science Communication.

2.3 Interpersonal deliberation as a debiasing strategy

The main aim is to test the potential for interpersonal deliberation in diverse groups to counteract the effects of identity-protective cognition.

Background. Due to the deleterious effects of identity-protective cognition on the accuracy of public perceptions of facts, much scholarly attention has been given to possible techniques of ‘debiasing’ (Lilienfeld et al., 2009). Examples include fostering a motivation for accuracy (Prior, Sood, & Khanna, 2015), instructing people to ‘consider the opposite’ (Lord, Lepper, & Preston, 1984), and decreasing identity threat through ‘self-affirmation’ (Sherman & Cohen, 2002).

One otherwise promising debiasing strategy that has received little attention in the context of identity-protective cognition is collective deliberation. For several domains, including deductive and inductive reasoning, mathematical problem solving and creativity, group performance has been found to be superior to individual performance (Moshman & Geil, 1998; Michaelsen et al., 1989; Laughlin & Ellis, 1986). One key factor in predicting group success is diversity: A variety of views within the group means that arguments for different perspectives are brought forth and tested, increasing the probability that a correct

solution is discovered (Schultz-Hardt et al., 2006; Watson et al., 2016; Yaniv, 2011; Krause et al., 2011). Nevertheless, no existing studies focusing on group performance has explored the impact of identity-protective cognition. It is therefore unclear whether identity-protective cognition will preclude the typical benefits of a diverse group setting, or whether such a setting can instead mitigate the impact of identity-protective cognition.

We will examine whether diverse groups that evaluate evidence about a culturally contested technology will show one striking effect of identity-protective cognition, namely attitude polarization – the tendency for partisans on an issue to strengthen their opinions of opposite directions following exposure to the same evidence (Lord, Ross, & Lepper, 1979; Miller et al., 1993; Cook & Lewandowsky, 2016 – or whether the presence of multiple perspectives mean that biases are effectively ‘neutralized’ - and likewise whether culturally homogeneous groups are particularly vulnerable to bias and resulting polarization.

Hypothesis.

- (5) Attitude polarization is reduced by interpersonal deliberation in culturally diverse groups but magnified by deliberation in homogenous groups.

Methods. We will conduct a laboratory experiment with individuals and groups (N≈404 - exact number depending on power calculation derived from survey results). Prior to arrival at the experiment proper, all participants will provide demographic information and complete cultural scales as identified as most informative in pilot testing and in the survey experiment. On this basis, we will quasi-randomly assign ~100 participants in the individual condition, ~152 will be assigned to culturally heterogeneous groups of four, and ~152 to culturally homogenous groups of four.

Upon arrival, all participants will individually indicate ethical attitudes and factual beliefs toward a target biotechnology (chosen on the basis of pilot testing and/or survey results). They will then assess two empirical studies pertaining to the technology, one suggesting its safety/benefit, the other suggesting its risk/lack of benefit. They will rate each piece of evidence for convincingness and once again indicate their ethical attitudes and factual beliefs about the technology. Participants in the group conditions will then be seated group-wise and deliberate about the evidence, while individuals engage in a thought-listing task (to control for time/effort differences). Following deliberation (and thought-listing), participants individually rate evidence for convincingness again and provide a final rating of ethical attitudes and factual beliefs. To identify biased assimilation, we compare differences in initial ratings of convincingness between cultural groups, and examine whether group context and group type changes these differences. To identify attitude polarization, we compare the difference between the first and second indication of attitudes/beliefs by cultural group and experimental condition.

Expected outcome

Paper 2.4: Biased assimilation and attitude polarization in groups: The role of cultural diversity.

Lead author: Bjørn Hallsson (postdoc). Publication venues: Psychology or communication journals, e.g. *Thinking & Reasoning*, *Journal of Personality and Social Psychology*.

2.4 Communication

The aim is to identify methods of communication that can neutralize the impact of identity-protective cognition on assessments of facts related to biotechnologies / assessments of risks related to biotechnologies.

Background. Identity-protective cognition poses a significant challenge to communicators (Sapp et al., 2013; Hart & Nisbet, 2012; Kahan, 2014a). Culturally motivated opponents of a technology may be very hard to persuade even on a solid basis of facts. Indeed, factual corrections of misperceptions surprisingly often fail (Nyhan & Reifler, 2010); and persuasive messages may result in ‘backfire’ or ‘boomerang’ effects, increasing opposition rather than decreasing it (Nyhan et al., 2014; Wood & Porter, 2016).

As mentioned in (2.3), several methods of decreasing the impact of identity-protective cognition on the evaluation of evidence have been proposed and tested. One promising venue is self-affirmation, a manipulation intended to affirm the value of an aspect of a person’s identity, with the aim of lowering the person’s ‘cognitive defenses’ (Cohen, Aronson & Steele, 2000). Self-affirmation has previously been shown to make subjects more objective when evaluating arguments about a personally relevant issue (Correll, Spencer & Zanna, 2004); to increase the amount of concessions in partisan group negotiations (Cohen et al., 2007); and to make people more likely to change their mind in response to threatening information (Sherman & Cohen, 2002). To the extent that self-affirmation can be integrated into communicative messages on a broader scale, it thus has potential as method of decreasing the impact of identity-protective reasoning on public attitudes.

While the effects of self-affirmation on processing of verbal arguments has been relatively well-documented, the ability of self-affirmation to protect against mistakes in judgments about logical or mathematical evidence has received less attention (although see Munro & Stansbury, 2009). This aspect is interesting from our vantage point, since identity-protective cognition does cause such errors (e.g. Kahan et al., 2017), which may be a substantial contributor to mistaken factual beliefs.

Hypothesis

(6) Self-affirmed participants will be more likely to correctly interpret mathematical evidence that threatens their worldview.

Methods. We will embed an experiment within a second nationally representative survey ($N \approx 2000$). In addition to demographic and cultural scale-items, we will measure subjects' cognitive ability (using the same measures as cited in (2.1)). Subjects in the self-affirmation condition will conduct an affirming task (see McQueen & Klein, 2006), while subjects in the no-affirming condition will conduct a similar, but non-affirming, control task. All subjects will then be asked to evaluate the results of a (fictional) study purporting to test the effects of a culturally contested technology (provisionally GMO or HPV vaccine) or a neutral technology (control condition). This evaluation will amount to a test of correlation detection; the ability to correctly infer a correlation on the basis of a 2×2 matrix of numbers (Munro & Stansbury, 2009; Kahan et al., 2017). Subjects will be asked which of 2 conclusions the study supports, one conclusion being the correct interpretation of the data. In the experimental condition, we vary whether the correct interpretation of the data is supportive of or threatening to the subjects' cultural worldviews. In the control condition, we expect cognitive ability to predict correct performance. In the experimental condition, we expect subjects to answer along cultural lines, and for cognitive ability to predict polarization more so than accuracy. However, we also expect that self-affirmation will diminish the effect of identity-protective cognition, such that cognitive ability increasingly predicts accuracy.

Expected outcome

Paper 2.5: Does Self-affirmation increase accuracy of mathematical reasoning about threatening information?

Lead author: Bjørn Hallsson (postdoc). Publication venues: Psychology or communication journals, e.g. Science Communication, Social Psychology, Journal of Communication.

WP3: Democratic legitimacy and the ethics of controversy

WP3 investigates how a democratic society ought to make decisions and formulate policies concerning NBTs, given that these are (or may become) controversial, and given that we have plausible psychological explanations of why they are controversial. In other words: How can we make democratically acceptable decisions in a domain where there are deep factual and normative disagreements between different members of the public, or between

the public and experts, especially when these disagreements are explained by identity-protective cognition?

While we will primarily be concerned with democratic governance of NBTs, our discussions will also shed light on issues that pertain to democratic theory more generally, since phenomena very similar to those characteristic of identity-protective cognition have been identified by political scientists with respect to people's political beliefs and preferences more generally (see e.g. Achen & Bartels, 2006; 2016; Campbell et al., 1960; Gerber & Huber, 2010; Gaines et al., 2007; Goren, 2005). WP3 will explore four aspects of democratically legitimate decision-making made salient by identity-protective cognition regarding controversial NBTs:

- (1) The common idea, widely implemented in science policy, that the public should participate directly in policy-making concerning NBTs.
- (2) The democratic relevance of citizens' views when these are shaped by identity-protective cognition.
- (3) The appropriate role of expertise and competent factual beliefs in democratically legitimate decision-making.
- (4) How established forums for dissemination of scientific knowledge and for conducting public debate can overcome identity-protective cognition and its undesirable polarizing effects.

3.1 Public participation in decision-making

Public engagement and participation in decision-making concerning NBTs is widely held to be an important goal (cf. Section 2 above). This sub-part of WP3 will critically examine the public participation ideal – i.e. whether there is a strong rationale for requiring public participation and what forms such participation should take.

3.1.1 Arguments for public participation

We will critically examine arguments for public participation as they appear in policy documents and in the scholarly literature. We expect to find two kinds of argument to be especially prominent. The first is an essentially negative argument against alternative ideas and approaches, especially the so-called deficit model of public understanding of science and technology, and technocratic approaches to risk management (Durant, 1999; Fiorino, 1990; Stirling, 2008). These approaches tend to privilege the expertise and perspective of technical experts, and public participation is argued to be required because it contributes important inputs that experts do not, such as non-technical expertise and 'public values'. The second argument holds that theories of democratic legitimacy – in particular deliberative proceduralist accounts – imply that public participation is required. We will examine these

arguments in light of theories of the respective roles of the public and of technical experts in deliberative democracy (e.g. Chambers, 2017; Christiano, 2012; Holst & Molander, 2017; Turner, 2013) and on the use of ‘mini-publics’ (e.g. Goodin, 2008; Fishkin, 2009; Parkinson, 2006). Our hypothesis is that the kinds of public participation envisaged by policy practitioners and stakeholders frequently go beyond what is required (and sometimes also beyond what is permitted) by the aim of legitimacy, including liberal requirements such as state neutrality, and that they will be likely to generate the wrong kind of input to the policy-making process due to the influence of identity-protective cognition.

3.1.2 How to structure participation

The theories of legitimacy mentioned imply a certain set of fairly specific goals for public participation exercises. These include, *inter alia*, constructive reason giving, restraint with respect to the reasons offered, and respect for scientifically established facts. The theory of identity-protective cognition suggests that the likelihood achieving these goals depends heavily on how deliberation is structured. Our hypothesis is that ensuring desirable results from deliberative fora requires more structuring than is commonly assumed, and will likely require the exclusion of advocacy groups that currently play a prominent role. Furthermore, it is likely that using formats of public participation that are not sufficiently structured will worsen the effects of identity-protective cognition and thus be counterproductive relative to the goals envisaged. We will critically evaluate existing formats of public participation that our literature review has revealed to be especially prominent, with the goal of suggesting reforms of these formats (or perhaps entirely new formats).

Expected outcome

Paper 3.1: The democratic legitimacy of public participation in policy-making concerning new biotechnologies

Paper 3.2: Identity-protective reasoning and public deliberation: Can we avoid the pitfalls?

Main author: Andreas Christiansen (postdoc). Publication venues: Science and technology studies journals and political philosophy journals, e.g. *Science, Technology and Human Values*, *Res Publica*.

3.2 Identity-protective cognition and democratic respect

Do policy preferences influenced by identity-protective cognition merit democratic respect? That is, suppose a citizen expresses a policy preference (e.g. that an NBT should be banned) and also holds false factual policy-relevant beliefs (e.g. that the NBT is highly risky) due to identity-protective cognition: Ought a democratically respectable state take

such a policy preference into account, or would it be legitimate to disregard it? A common view assumes that (i) citizens' views earn their democratic respectability by constituting some kind of value judgment (e.g. concerning interests, or reasons, or principles), with the important qualification that the judgment must be reasonable, or must have or be able to survive critical scrutiny, and (ii) that where factual beliefs are concerned, truth or epistemic justification takes precedence over democratic respect for the views of each. This view is in line with the orthodox conception of practical reasoning, which assumes that people's basic values or aims combine with their factual beliefs to produce preferences for policies that are congruent with one's values, and that further one's aims.

An underlying assumption of the view just described is that factual beliefs and values judgments are generated independently of one another and answer to different criteria of justification: Factual beliefs answer to the evidence, while values answer to the norms of practical reasoning. However, the psychological evidence in favor of identity-protective cognition suggests that agents with our actual psychology tend to exhibit a more complicated relationship between factual beliefs, attachments to identity groups, values, and policy preferences: Political-cultural worldviews are expressions of fundamental political values and attachments to identity-groups, and they strongly influence people's factual beliefs. This presents a puzzle concerning the democratic respectability of citizens' preferences. A common approach has been to assign democratic respectability to informed and/or rational preferences (e.g. Gaus, 2011; Goodin, 1995, ch. 9; Nagel, 1987; Rawls, 1993; Vallier & D'Agostino, 2014). Identity-protective cognition raises a puzzle for this kind of approach. Identity-protective cognition may cause a citizen S to hold false beliefs on matters of fact relevant to a policy area in which S holds the policy preference P. On one conception of informed preference, P is not an informed preference, since S does not hold true policy-relevant beliefs. However, on a counterfactual conception of informed preference, P could well be an informed preference. On such a conception, P is an informed preference only if S would continue to hold P, were S to hold true policy-relevant beliefs. Since S's false beliefs are caused by her fundamental political values, S might well retain P on the basis of those values alone. So because S's factual beliefs express S's political views, their falsity does not undermine their democratic respectability (cf. Kahan et al., 2006; Kahan & Braman, 2007).

It is thus unclear whether policy preferences are democratically respectable when citizens hold false policy-relevant beliefs due to identity-protective cognition. We suspect that this fact reveals a general weakness of the informed-preference strategy. We will investigate how (if at all) such a strategy can be made to work. We will construct a set of normative criteria for democratic respectability that are influential in democratic theory and test whether any specification of informed preferences are able to meet those criteria in light of the psychological facts.

Expected outcome

Paper 3.3: Democratic responsiveness and identity-protective reasoning

Lead author: Andreas Christiansen (postdoc). Publication venue: Political philosophy journals, e.g. Journal of Political Philosophy.

3.3 Democratic legitimacy and competent factual belief

The fact that citizens often lack well-informed opinions on even highly salient policy issues is nothing new (Converse, 1964; Kuklinski & Peyton, 2007; Zaller, 1992). Philosophical discussions of the proper relationship between competence in factual belief and the legitimacy of democratic procedures are likewise not new, and the issue remains hotly debated. We will contribute to this debate in three ways:

(i) One widely discussed proposal in this area is the institution of an *epistocracy*. In an epistocracy, those who are better informed and more competent are given extra political influence in some way. Concrete proposals include voter examinations, such that the right to vote is conditional on passing a test of “generally relevant social science and basic knowledge about the candidates” with the aim of “screening out citizens who are badly misinformed or ignorant” (Brennan, 2011, p. 714), and giving extra votes to those who are better educated (Mill, 1861, ch. 8). The fact that better informed and more educated people are more likely to hold false beliefs, as identity-protective cognition theory shows, seems to undermine the case for these sorts of moderate epistocratic schemes. We will develop this insight into an argument against epistocratic schemes of the kinds that have been suggested.

(ii) As noted, the case of risk management has played an especially central role in the argument for public participation. The proper role of experts in risk management has similarly been subject to much debate. On the one hand, risk management obviously requires complicated technical knowledge about which risks are serious and how they can be mitigated (Margolis, 1996; Sunstein, 2002; 2005). On the other hand, the delegation of authority to experts has been criticized on democratic grounds. For example, it has been argued that the procedures utilized by experts, most prominently risk-cost-benefit analysis, ignore legitimate value dimensions in risk management to which lay people are sensitive (Slovic, 2000; 2010). We will investigate the merits of this democratic criticism of expert decision-making in the case of risk. We will ask (a) whether (and to what extent) the democratic criticism is warranted, and (b) whether and how expert procedures can be reformed so as to become democratically legitimate.

(iii) It is highly intuitively plausible that factual incompetence is in some way a political problem. For example, Thomas Christiano argues that democratic procedures should be “truth sensitive” (Christiano, 2012); Philip Deen argues that “political authority is, in some way, grounded in the truth” (Deen, 2014); and Jason Brennan argues that it is unjust to impose coercive laws on people “as a result of decisions made in an incompetent ... way” (Brennan, 2011). However, some foundational questions remain unclear, in particular (a) what is the nature of the requirement in question (e.g. is it truth-based or competency-based?), and (b) what is the precise relationship between policy-relevant facts and legitimate policy (e.g. how does the fact that there anthropogenic climate change is occurring constrain legitimate policies?). We will attempt to clarify these foundational issues and thus come to a better understanding of what role factual competence has to play in a theory of democratic legitimacy.

Expected outcome

Paper 3.4: The inverse relationship between information and true belief: A new argument against epistocracy

Paper 3.5: Experts, citizens and democratic legitimacy in risk management

Paper 3.6: How do facts and factual competence matter for legitimacy?

Lead author: Andreas Christiansen (postdoc). Publication venue: Political philosophy and social epistemology journals, e.g. *Episteme*, *Politics Philosophy and Economics*, *Philosophy and Public Affairs*.

3.4 Improving the democratic conversation

A broad-based and civil democratic conversation remains an important aim, also with respect to issues of science and technology. Such a conversation depends on the existence of a “common reality” on which we agree, which acts as a basis and anchor of discussions, and which should preferably include the facts (Lynch, 2017). This, essentially, is what is lacking when politics become ‘post-factual’. The existence of this common reality further rests on a number of institutions that produce and disseminate factual information, and on norms constraining participants in public debates in ways that ensure a productive and open discussion.

The empirical work we conduct in WP2 will generate insights into how this common reality may be threatened by identity-protective cognition, and how the threat can be averted. We will translate these insights into recommendations for how the institutions and norms on which the democratic conversation depends should be reformed. The output of this work will consist partly in efforts to influence how the institutions and actors that shape the democratic conversation think and behave, e.g. through writing articles for newspapers

and trade magazines (described in the dissemination plan below), and partly in contributions to scholarly journals in political philosophy and science communication. With respect to political philosophy, we will target the debate on reasonability in public debate. On the mainstream (“consensus”) view of public justification, reasonability requires us to only offer reasons that we sincerely believe others can accept as reasons (Vallier & D’Agostino, 2014, §2.3). The existence of psychological biases, including those associated with identity-protective reasoning, poses challenges for this ideal, since we can too easily deceive ourselves into believing we are more reasonable than we really are (Kahan, 2007; Lynch, 2016). We will suggest criteria of self-criticism and restraint concerning factual beliefs that improve our chances of being reasonable. With respect to science communication, our working assumption is that political neutrality and the creation of a better understanding of the facts are important goals of science communication. We will suggest how these goals can better be achieved by being cognizant of the triggers of identity-protective reasoning.

Expected outcome

Paper 3.7: Identity-protective reasoning and demands of reasonableness in public deliberation: Self-criticism and appropriate factual restraint

Paper 3.8: Identity-protective reasoning and public understanding of science: How can science communicators avoid politicizing new technologies?

Lead author: Andreas Christiansen (postdoc). Publication venues: Political philosophy and science and technology studies journals, e.g. *Public Affairs Quarterly*, *Public Understanding of Science*.

WP4: Using the convergence framework for systematic ethical reviews of selected NBTs

In this work package we will perform systematic ethical reviews of a select range of appropriately specific applications of NBTs. The reviews, and the selection of their precise targets, will be done in collaboration with leading scientists from biomedical and biological research communities that are part of the project, ensuring that analyses combine leading expertise on the scientific aspects of a technology with the group’s expertise on ethics. The process, format and overall analytical strategy for these reviews will be adapted from a process used in earlier work (Andersen et al., 2015; Palmgren et al., 2015; Østerberg et al., 2017): First, in collaboration with our partners, we decide for an appropriately specific application of an NBT that will be the object of analysis. The application should be scientifically realistic, and could be either an existing one or one that may potentially come into use. Second, it is described in detail how the application works in realistic scenarios. Third,

we describe (briefly) the legal, institutional and economic setting in which the target application works or is assumed to work. Fourth, we identify the range of ethical objections, concerns and worries that have been expressed in the academic literature and in public opinion, if evidence concerning this exists. Fifth, we analyze these concerns in the light of convergence framework established in WP1, and in terms of the general features of identity-protective reasoning (WP2) and the political philosophy of ethical controversy (WP3). In general, this will require analytically determining whether the ethical concerns are consistent with the convergence framework or not. If an expressed concern is consistent with the framework, then we try to determine the implication for the evaluation of the application in question. If an objection is not consistent with the convergence framework, we try to determine the plausibility of the objection (are there independent weighty philosophical reasons supporting it?), and the implication for the target application. As discussed in WP3, in some cases non-standard objections may be set aside reasons of political legitimacy. In addition, we will write a paper that presents and critically evaluates the convergence framework for ethical analysis of NBTs employed. The expected outcome of WP4 will be:

Paper 4.1 on gene editing (with professor Hans Wandall)

Paper 4.2 on novel plant biotechnology (with professor Michael Palmgreen)

Paper 4.3 on clinical genetics (with professor Elsebet Østergaard)

Paper 4.4 on personalized medicine (with professor Søren Brunak)

Paper 4.5 on assisted reproduction (with professor Claus Yding Andersen)

Paper 4.6 on stem cells (with professor Thomas G. Jensen)

Paper 4.7 a general presentation and evaluation of the convergence framework for ethical analysis of NBTs.

Main author of these papers will be Klemens Kappel, in collaboration with the above. The review papers will be published in journals in the relevant scientific fields (e.g. Cell, Trends in Plant Science). The general presentation and evaluation of the convergence framework will be published in multidisciplinary aimed journals (e.g. Journal of Medical Ethics, Bioethics), and/or general science journals (e.g. Science, Nature Biotechnology).

4. Dissemination plan

The scientific outcomes of the project will be disseminated through publications in broad and specialized academic journals, through conference participation and through a series of workshops that we will host. In addition, we will execute a comprehensive dissemination effort targeted at the Danish public opinion, decision-makers, scientists, and science communication professionals. The details of the dissemination plan covers a number of differ-

ent target audiences, avenues for dissemination and specific actions to be undertaken (where some actions covers several audiences and key avenues):

Target audience	Key avenues for dissemination	Specific actions to be undertaken
Academics in philosophy, psychology, and science and technology studies	Scientific journals in philosophy, psychology and STS, academic conferences, workshops	(A) academic publications (B) presentations at workshops and conferences
Researchers in bio-medical and plant sciences	Publications in science/tech journals, presentations and talks, interaction with research groups	(C) Publications in specialized journals (D) Roadshow for biotech researchers (E) Book in Danish for general audience
Policy makers	Interventions in the ordinary public debate in newspapers, TV, blogs; town hall meetings, public hearings etc.	(E) (F) Ad hoc interventions in public debate (G) Appearance at Folkemødet 2019
Professionals working in media and science communication	Articles written in the general media as well as in communication-specific publications.	(E) (F) (H) Paper in communication-specific publication, workshop for this audience
High school students	Appearances at high schools and high school talent programs.	(E) (I) Roadshow for high school students
Sub-groups of the general population interested in science, technology and its societal aspects	Appearances at science festivals such as Bloom, Kulturnat, Forsknings Døgn; articles in Danish newspapers and outlets featuring science news.	(E) (J) Appearance at Bloom Science festival (K) Dissemination papers for Weekendavisen's science section, videnskab.dk, Ingeniøren, Videnskab.dk, and similar
The public at large	Interventions in general public debates that may arise where our expertise is relevant (e.g. vaccine skepticism, GMO, gene therapy).	(E) (F)

As for (A), (B) and (C), see above and Section 5. For more timeline of deliverables, see the overall project timeline, Section 7. Details of other actions are:

(D) Roadshow for biotech researchers in DK. We develop a roadshow presenting main findings in cognitive psychology and discussions of implications for ethics, democracy and science communication targeted for biotech researchers. The show will be presented to our partner's research groups, and to other interested parties, e.g. "Akademiet for Talentfulde Unge", for specialized scientific societies or seminars for young researchers (e.g. the "Forskospirer" program). Development of the roadshow will start in Year 1, Q1.

(E) Book in Danish for a general audience. The book should be short (100-150 pages), and easily accessible, and published with a main publishing house in Denmark, e.g. Gyldendal. The book will outline the main findings in cognitive psychology on identity-protective reasoning, and discuss the ethical implication for science-controversies, science communication, media coverage and public deliberation and democracy in general. The book will be collaboratively written with Klemens Kappel as lead author. Work on the book will start Year 1, Q1 where we will draft an outline and contact a publisher. The book will be collaboratively written over the next three quarters. Beginning the book project early will promote collaboration and cohesiveness, generate ideas for further research, and the published book will provide a vehicle for public attention.

(F) Ad hoc interventions in public debates in Denmark. We will talk to journalists and participate in various public debates in radio and television on an ad hoc basis.

(G) Appearance at Folkemødet 2019 (a major annual public event for policy-makers, media professionals and the general public). We will work with Faculty of Humanities, University of Copenhagen, Danish Ethical Council or other interested organizations to set up talks/panels with topics related to the project, for example on political polarization and science-controversies, and implication for science communication, public deliberation and democracy in general. Planning our appearance at Folkemødet 2019 will start in Year 1, Q3.

(H) Dissemination to science-communication practitioners in DK. We will write a paper on the general challenges involved in identity-protective reasoning of the project for a specific outlet for science communication practitioners (e.g. Kommunikationsforum) (Year 2, Q1). At the end of the project we will invite science communication practitioners to a workshop discussing the main findings of the project, or present work on the bi-annual conference for Danish science journalists.

(I) Roadshow for high school students in Denmark. We will adapt a version of the roadshow targeted to high school students, especially those taking the 'biotechnology' elective

class, presenting main findings in cognitive psychology and possible implications for ethics, democracy and science communication. To be presented for high schools around the country. The adaptation will be produced in Year 2, Q1.

(J) Appearance at Bloom science festival 2018. We will organize a panel at the annual nature and science festival *Bloom*, which is to be held in Copenhagen, May 27th-28th 2018. Potential topics include psychological understandings of risk perception, public attitudes to science and technology, the science of science communication, implications of identity-protective reasoning for public deliberation and democracy. Planning will begin in Year 1, Q1.

(K) Dissemination papers for *Weekendavisen's* science section, *videnskab.dk*, *Ingeniøren*, and similar. These will be short papers and articles. All members of the project will be engaged in this. We will continuously consider whether we have findings or materials that fit this purpose.

5. Interaction with existing research environments and experts

Collaboration with NBT science community in DK. The project involves close collaboration with Professor Søren Brunak (personalized medicine / Novo Nordisk Foundation Center for Protein Research, UCPH), Professor Thomas G. Jensen (stem cells / Dept. of Biomedicine, AU), Professor Michael Palmgreen (Plants / PLEN, UCPH), Professor Hans Wandall (gene editing / Centre for Glycomics, UCPH), Claus Yding Andersen (reproductive medicine / Rigshospitalet), and Elsebet Østergaard (clinical genetics / Rigshospitalet). The collaboration will consist in writing jointly authored papers (see WP4 and project timeline). Moreover, all NBT science collaborators are also part of the project's advisory board (see section 6).

International panel and national collaboration. The group will collaborate with a panel of leading international researchers, research groups and centers within relevant areas of philosophy and psychology, including professor Thomas Douglas (the Uehiro Centre for Practical Ethics, University of Oxford), professor Michael Lynch (Humanities Institute, University of Connecticut), professor Robert Talisse (Philosophy, Vanderbilt University), Pascal Borry (Centre for Biomedical Ethics and Law, Leuven University), Heidi C. Howard (Centre for Research Ethics and Bioethics, Uppsala University). The venue for the international collaboration will be a series of workshops that we host. Workshops are typically 2-day events featuring 2-4 international scholars working in areas of interest to the project. We plan to host one such workshop pr. semester. Workshop participants will include our

international collaborators, and other scholars working in these areas. Members of the CEEC Research Group will present work in progress, and the workshops will thus provide a highly valuable venue for younger scholars to discuss their work with international scholars, will ensure that our work is informed by latest developments in international research, and will secure impact on the international research agenda. In addition, we will arrange for short research visits for members of the research group to research environments abroad. The group has excellent contacts with other parts of the Danish ethics research environments through and they will also be invited to participate in research activities and to present papers in our international workshops. Our national collaborating partners include Professor Jesper Ryberg (RUC), associate professor Martin Marchman Andersen (DTU), and professor Mette Nordahl Svendsen (Department of Public Health, KU (the MeInW project)).

The CEEC research group and the project will be located at the Section of Philosophy, Institute for Media, Cognition and Communication, University of Copenhagen, and the institute will provide office space and full administrative support to the group. The proposal brings forward excellent parts of three different projects under the UCPH Excellence Program for Interdisciplinary Research namely “Plants for a Changing World” (Co-PI: Klemens Kappel), “bioSYNergy” (Co-PI: Sune Holm), and “Global Genes, Local Concerns” (Co-PI: Klemens Kappel), as well as the FKK-funded project “Complex Disagreement” (PI: Klemens Kappel). The three post-docs have all contributed to these projects (Nana Kongsholm has done her PhD in “Global Genes, Local Concerns”; Bjørn Hallsson has done his PhD in “Complex Disagreement”; and Andreas Christiansen has done his PhD in “bioSYNergy” and a postdoc in “Plants for a Changing World”). The Section of Philosophy at University of Copenhagen is home to a number of large international research projects in philosophy, and is ranked 29 best in the world (QS World University Ranking 2017).

6. Organization and governance structure

CEEC Research Group. PI of the project will be professor Klemens Kappel, who will lead the CEEC Research Group consisting of Bjørn Hallsson (postdoc), Nana Cecilie Halmsted Kongsholm (postdoc), Andreas Christiansen (postdoc), Katla Heðinsdóttir (PhD-student), and the research assistants working in the project. The group will also include associate professor Sune Holm and associate professor Morten Ebbe Juul Nielsen, who both have international research expertise in related areas. The CEEC Research Group will meet weekly to present and discuss the group’s work in progress, papers of joint interest, and to address issues of strategy, planning, workshops, dissemination initiatives, and bids for further funding. The mode of collaboration in the group will support and encourage

collaborative work, which is a mode of working that all members are trained in and familiar with. All members of the group will be located in the Section of Philosophy, University of Copenhagen, and will work there on a daily basis. Klemens Kappel, who has directed several similar research groups and large research projects, will assume overall managerial responsibility for planning and executing all activities in the project. The postdocs in the project will, under the supervision of KK, act as Co-PIs for some of the work packages: WP1: Nana Kongsholm, WP2: Bjørn Hallsson, WP3: Andreas Christiansen. Klemens Kappel will direct WP4 and the dissemination effort.

Advisory Board. The project will be overseen by an advisory board consisting of Professor Maja Horst (science communication scholar, Head of Institute of Media, Cognition and Communication, University of Copenhagen), professor Jesper Ryberg (Philosophy, Roskilde University), professor Søren Brunak (Novo Nordisk Foundation Center for Protein Research, University of

Copenhagen), professor Thomas G. Jensen (Head of Dept. of Biomedicine, Aarhus University), professor Michael Palmgreen (Dept. of Plant and Environmental Science, University of Copenhagen), professor Hans Wandall (Centre for Glycomics, University of Copenhagen), professor Claus Yding Andersen, (reproductive medicine, Rigshospitalet), PhD Morten Andreasen (academic staff, Danish Ethical Council), Elsebet Østergaard, (Clinical Genetics, Rigshospitalet). The role of the advisory board will be to discuss the general issues pertaining to the project, the developments of the project, secure embedment in science and research, help facilitate dissemination, and propose directions of future research. The Advisory Board will meet with the PI and the research group once every 6 months.

