Imagine a patient, Bill, who has lost decision-making capacity. An important treatment decision must be made, but Bill cannot make it. Assume there is no advance directive or anything equivalent. There is no way to take a statement made by Bill, when competent, and apply it to the current setting. A standard move in cases like this is to have the patient’s surrogate decision-maker—Bill’s life-partner, for example—enter the picture and answer the question, “In this situation, what would Bill choose?” Imagine that Bill is in his late sixties and has always prided himself on his mental acuity, that he has held a job with considerable decision-making responsibility and in his spare time would read voraciously and do crosswords and acrostics. The proposed lifesaving procedure has a 70 percent chance of leaving Bill with significantly decreased cognitive function. Bill’s partner believes that she knows what Bill would choose: Bill would not want to live with severely limited mental capacities, and he would decline a procedure likely to leave him in this condition. The medical team is ready to accept her decision. Something seems very right about taking this route, but what makes this the right thing to do?

Bioethics orthodoxy sees the patient’s autonomy as at stake here, but in fact this is not so, at least not as “autonomy” is usually understood in clinical contexts. Instead, what underpins the moral force of the “What would the patient choose?” question is a value that is related to—but not the same as—either autonomy or that other common bioethical consideration, the patient’s best interests. Or so I will argue.

I hope that clinicians will find little new in what I say. I suspect that most of them already use the concept I will introduce. Actually, my goal is simply to make transparent what I have heard discussed during five years of weekly case conference meetings at the MacLean Center for Clinical Medical Ethics at the University of Chicago. Those discussions have been nuanced in a way that fits poorly with the limited

An Outline of Surrogate Decision-Making

When a patient is declared not competent and a treatment decision must be made, the patient’s surrogate is supposed to step in. On the standard view, the surrogate should proceed as follows:¹

- If the patient has made out an advance directive or has in some other way made his wishes clear, the surrogate should decide in accordance with those wishes. She should do what the patient indicated he wants done.

- If there is no sufficiently clear statement of the patient’s wishes, the surrogate should try to answer the question, “What would the patient choose to do if he could be made aware of the choice in front of him?” In trying to answer this question, the surrogate is supposed to rely on her knowledge of the patient’s beliefs and values. She is supposed to figure out what his choice would be.

- If the surrogate doesn’t have sufficient knowledge to answer this question, she is supposed to decide on the basis of her beliefs about where the patient’s best interests lie.

The patient’s choice is thus relevant at two stages. If he made out an advance directive or in some other way said what he wanted, his actual choice is relevant. And if there is no sufficiently clear expression of his actual choice, the surrogate should try to determine his hypothetical choice—what, if competent, he would choose.

Self-Determination and Authenticity

To get clear on the issues in surrogate decision-making, we need to think about which human capacities are relevant to the surrogate context. To begin with, humans have the capacity to choose, to take something as a sufficient reason to act. Kant calls this our capacity “to set ends.”² Sometimes it is talked of as the “capacity for rational self-rule.”³ I will call it the capacity for self-determination. It is the capacity to make a choice by assessing the pros and cons of competing alternatives.

Beyond merely choosing, we also have the capacity to be a particular self, a distinctive individual—in the words of the Sinatra song, to do it “my way.” This ideal goes back to the Renaissance but receives its canonical form in John Stuart Mill’s On Liberty.⁴ I will call this value authenticity.

Finally, we have the capacity to live a good life, to have a set of interests that can be satisfied in ways that are better or worse—to have best interests.

Each of these capacities encapsulates an important value, and these bear interesting relations to one another. To start with, a person’s best interests are not identical to either self-determination or authenticity. Both are elements in her best interests: she has an interest in making decisions and an interest in expressing her individuality. But she also has an interest in leading a life that is good, whether that involves helping others or realizing her artistic capacities or something else. The goodness of a person’s life—what makes it a good life—inherently involves more than self-determination and authenticity.

Now, self-determination is not the same as authenticity. Nevertheless, the two are often evidence for one another’s presence. In wondering whether to accept a patient’s decision, the medical team might ask if the decision is authentic, meaning: Does it fit with her prior beliefs and values? Should it really count as “hers”? Self-determination is likely not to be present if the content of the patient’s choice is far out of line with who she is. Significant deviation from her usual beliefs and values is reason to worry that her capacity to make choices has gone awry.⁶ Conversely, if we have reason to think that the patient’s self-determination might be jeopardized—as might be the case if she is septic, for example—then we might also wonder whether her current choices are authentic.

That either capacity is robust, then, is reason to be confident that the other is also functioning well. Still, each is intelligible apart from the other. It might happen that a man has never said anything about how to hold his funeral, but that we believe his funeral should include a chorus of Harley-Davidson engines on the grounds that it fits our understanding of who the man was. There is no issue of self-determination here—of the man’s decision-making—but instead a conception of his beliefs and values and of their implications for his funeral. Of course, people sometimes act out of character, and while this might prompt a doubt about whether there really was self-determination, the doubt can usually be allayed. Part of why deciding for oneself is valuable is that those decisions cumulatively make one the self one is (self-determination is, in a sense, the ongoing determination of the self). Still, most of us value deciding for ourselves even apart from the impact of such decisions in constituting the self. Most of us want to be treated as a being who can, at this very moment, make up her own mind.⁷

Bioethics orthodoxy holds that the central moral issues in surrogate decision-making are the patient’s autono-
my and best interests. However, I think that when practicing physicians invoke the concept of autonomy, they usually have in mind both self-determination and authenticity. To go back to Sinatra’s song, what is thought to be important is that both I do it (self-determination) and that the way I do it is “my way” (authenticity). This conflation is natural enough. Making my own decisions is linked to living a life that is mine. So far as we wish to know what is valuable about autonomy, both elements could be invoked. In most actual decision-making, both are present. Clinical discussions frequently shuttle between the two. In deciding what the patient would choose, however, one of these elements drops out.

An Entrenched Conceptual Mistake

Let’s return to the path a surrogate is supposed to traverse. In practice, this path may diverge from the straight and narrow. A surrogate might make a decision that is inconsistent with an advance directive’s explicit content, and the physicians might go along with her. They might do so to avoid bedside conflict or for some other reason not related to the patient’s choice. However, there are also good reasons to worry about the relevance of the advance directive. Was what the patient decided earlier consistent with his current beliefs and values? Was it too vague or general to capture the specifics of the moment when a vital decision must be made? As for the second step on the surrogate’s path—the “What would the patient choose?” question—there is plenty of evidence that surrogates are inaccurate at predicting what patients would choose.

These are all serious issues, but I want to turn from them and focus instead on an ambiguity in the “What would the patient choose?” question. This ambiguity often leads to a philosophical mistake.

Imagine that last week, while fully competent and well-informed, Jane made out an advance directive saying that should situation S occur, she wants all treatment stopped. (She might also have indicated her wishes in some other way, but for simplicity’s sake, let’s suppose she used an advance directive.) This week, Jane is no longer competent and situation S has occurred. The current orthodoxy says that, on this description of the case, treatment should be stopped. But why is that the right thing to do?

The standard answer is that to fail to comply with Jane’s advance directive would be to override her autonomy and will. That is why they should be highly valued.

This might seem clear enough. Unfortunately, there can be an ambiguity about the surrogate’s job. Sometimes it is described as follows: Go over your knowledge of the patient in order to call up instances in which the patient expressed herself, directly or indirectly, about what she wants done in the current circumstances. Jane’s case exemplifies this process. Jane’s surrogate knows that Jane has expressed herself directly through her advance directive.

If we believe someone’s funeral should include a chorus of Harley-Davidson engines because it fits our understanding of who the man was, then there is no issue of self-determination at stake, but instead a conception of his beliefs and values and of their implications for his funeral.

However, sometimes the surrogate’s job is described differently: Go over your knowledge of the patient in order to figure out, from your grasp of the patients’ beliefs and values, what she would want done in the current circumstances. This is the surrogate’s job if there is no sufficiently clear statement of the patient’s wishes—in other words, no sufficiently clear prior exercise of the patient’s will—and in such a case the surrogate moves along the standard path to the second step. This step is exemplified by the decision-making for Bill, the patient we met at the beginning, who has never said what he wants done if he can’t make his own medical decisions. Fortunately, Bill’s surrogate has a good sense of Bill’s beliefs and values, and so she has a good sense of what Bill would want done. The point to keep in mind is that the surrogate’s task in deciding for Jane is very different from her task in deciding for Bill.
and the different tasks will be justified by different moral concepts.

With Jane, the focus is on what Jane actually expressed at some point in the past. It is irrelevant whether Jane used a formal advance directive or in some other way made her choice plain. The reason to defer to Jane’s choice, however expressed, is that to do so respects her capacity for self-determination. The difference between an advance directive and the surrogate’s memory of what a patient said is merely a question of evidence. With an advance directive, we are more confident that we are following the patient’s actually expressed choice. In other words, it concerns a hypothetical choice. So the moral basis here has nothing to do with respecting the patient’s capacity for self-determination. When we ask the question “What would the patient choose?” it might look as if we are engaged in trying to understand the content of the patient’s will. It might seem that if we could get a reliable answer to that question and then deferred to it, we would be giving effect to the patient’s will. We could then appeal to the great moral importance of self-determination to explain why deference should be accorded the answer to that question. In fact, appeal here to Bill’s capacity for self-determination is a conceptual error, for Bill no longer has that capacity.12

This error is one with which political philosophers have long been familiar. Hypothetical consent (for instance, in recent versions of social contract theory) is not a weaker form of actual consent. Rather, the two are conceptually distinct. Hypothetical consent has to do with what a person would agree to, not with what she has agreed to. Hypothetical consent has nothing at all to do with the exercise of a person’s will. It concerns what would be rational to do, not what is actually being done. In the social contract tradition, the appeal to the state of nature was originally an appeal to a supposedly real state of affairs and so to supposedly actual events—for example, to an actual exchange of promises.13 Now the moral relevance of a promise is that it is a specific kind of exercise of the will: one deliberately binds oneself to do something in the future. By contrast, there is nothing morally binding about a hypothetical contract. Invoking such a contract is a way to get at what action makes the most sense in a given situation. Here, “choice” is just a metaphor—a way to track what is reasonable to do. The will is not in the picture.14

This is equally true of the “What would the patient choose?” question. It, too, is about a merely hypothetical choice—something distinct from the exercise of the patient’s will. One can see this easily enough. The reason to find a surrogate for Bill is that here and now, at this very moment, Bill is not capable of the competent exercise of his will. He is no longer self-determining. By hypothesis, there can be no present competent exercise of Bill’s will; otherwise, he wouldn’t need a surrogate. It follows that if we move along the decision path to the point where we reach the second description of the surrogate’s job (as in Bill’s case), the patient’s will is no longer in play—neither his past nor his present will. That means that if we have reached this point, we cannot defer to the patient’s capacity for self-determination, either past or present. We are no longer in a position either to respect or to disrespect that capacity. We can neither uphold nor violate it. Thus, in Bill’s case, the moral value of self-determination cannot be the moral basis of the appeal to the question, “What would the patient choose?”

As so much clinical literature makes this conceptual mistake, I am going to belabor the point a moment.15 What is of value when we talk about self-determination is the exercise of the will. If the patient needs a surrogate, it is because he can no longer do so. If Bill needs a surrogate, then he is not capable of exercising his will. And if we have inadequate information about Bill’s past exercise of the will, then to defer to the answer to the “What would Bill choose?” question would not be somehow to exercise Bill’s will “for him.” That is, quite simply, something that cannot be done.16

I think the distinction is sufficiently important that we should ask different questions, corresponding to the different forms that surrogate contexts can take. With Jane’s case, there is a clear enough advance directive or other prior exercise of the patient’s will. Here, the proper question to ask the surrogate is not “What would the patient choose?” but “What did the patient choose?” With Bill’s case, there is no such clarity about the patient’s prior will. This is the place for the question, “What would the patient choose?”

The Complexity of Actual Cases

A persistent philosophical mistake tends to be well motivated. Most people in our culture have the intuition that, in surrogate contexts, the “What would the patient choose?” question is an important one, and its answer has considerable moral force. The question invokes the idea of choice, which often has moral force, and it is natural enough—even if here it is mistaken—to tie the value of choice to self-determination, and so to think that self-determination accounts for the moral force of the question’s answer.

The complexity of the surrogate’s job contributes to the confusion. There is, for instance, the problem of determining whether there was a prior exercise of the patient’s will, and even if there was, of determining its content. Suppose that Jane has stated in an advance directive that she should never be put on a ventilator.
Was her refusal supposed to extend to short-term ventilation during an acute illness? Merely having a piece of paper or even a tape of what Jane said might not be sufficient to answer such a question. As with Bill, it might be crucial for Jane's surrogate to understand her beliefs and values. Nevertheless, the issue is still whether Jane actually exercised her will in a way that covers the current context, and the job for the surrogate is to determine what Jane actually did, not to figure out what she would do.\(^{17}\)

However, matters might be even more complicated. It is not only that even explicit statements may need interpretation. Suppose that with another patient, Dave, the surrogate remembers something Dave said about not wanting treatment, and that either the memory is vague or Dave's comments were vague. In trying to dispel the vagueness, the surrogate might find herself thinking about what she knows of Dave's beliefs and values. She might not be confident that Dave ever directly expressed himself about treatment, and she might not be confident that she knows in a straightforward way how Dave's beliefs and values bear on treatment. Yet she might think that Dave said enough and that she knows enough that, on the whole, she knows what should be done. Asking the surrogate to consider both elements often makes good practical sense. Still, morally speaking, the two elements diverge. The first is a report on events that happened in the past; the second is an attempt to apply a set of beliefs and values.

**Authenticity and Best Interests**

In many situations, the question of what a person would choose has obvious importance. I am late to the restaurant and you have to order for me. If you want me to enjoy my meal, then ordering what I would choose makes good sense. Of course, the connection is contingent. Maybe I would find even more enjoyment in something that I'm too stuck in my ways to try, and this is a chance to expand my gustatory range. Or maybe my best interests go beyond my momentary pleasure, and since I would certainly choose the bacon-double-cheeseburger, you think my best interests require a preemptive ordering of the vegan special. What I would choose is no doubt a fallible guide to my best interests; still, it is a guide.

This is well and good. However, if we turn now to the value of authenticity, we will see that it points to a deeper thought about the “What would the patient choose?” question. At stake is often said to be the idea—indeed the ideal—of living a life that expresses my individuality. Many writers have stressed the importance of this ideal. It is connected to Ronald Dworkin’s insistence on a life’s “integrity” and to the thought, urged by many, that a life is really a narrative, that a more coherent narrative is a better narrative, and so a more coherent life is a better life.\(^{18}\) On this view, deferring to the answer to the “What would the patient choose?” question maintains the coherence of the patient’s life because the content of the patient’s hypothetical choice reflects the beliefs and values by which he has lived. The patient’s life would thus continue, or be brought to a close, in accordance with his beliefs and values, and that would make it a better life. The ideal of authenticity is a powerful one, with deep roots in our culture. I think it is the value of authenticity that gives the “What would the patient choose?” question its central moral force.

An objection could be raised here. Given that authenticity’s value is in making a person’s life better, strictly speaking it merely advances the agent’s best interests. So why not class it with those interests? Why think of it as a distinct value?

The reason is that authenticity is only one element in a person’s overall best interests and, though very important, it might nevertheless be less important than others. In fact, the patient’s overall best interests might be at odds with her authenticity. Keep in mind that, as shown by my penchant for bacon-double-cheeseburgers, a person’s best interests are not reducible to the satisfaction of her desires. Those desires might be problematic in various ways. They might be mutually inconsistent or based on false beliefs, poor reasoning, bad judgment, or pernicious values. Philosophers almost universally deny that the conditions for a good life involve merely satisfying the desires a person just happens to have.\(^{19}\) But to have my life go “my way” is precisely about satisfying my desires. The basic claim about “my way” is that it expresses me, this person, who has a particular set of beliefs and values, however faulty. Yet at some point my beliefs and values might become sufficiently faulty that doing it my way would make for a worse life, even if a more authentic one. At some point, then, authenticity and overall best interests might part company. Because

---

**Although authenticity is highly important, it isn’t a trump. In some cases, other values might override it. Most of the time, it will be the patient’s own best interests that compete with authenticity, but other interests, such as the family’s interests, might also be relevant.**
they might do so, keeping the two values distinct makes for both analytic and practical clarity.

To return again to Bill, it is the value of authenticity that would make not doing the offered procedure the right choice. In Bill’s case, his surrogate judges that, if asked, Bill would have refused the procedure because of the high risk of ending up with severe compromised cognitive functions—that is, of ending up with a life deeply out of keeping with what he has always valued. The right thing to do, his surrogate decides, is not to run the high risk that Bill would end up with a life that, in a basic sense, would not be his own.

The Limits of Authenticity

The upshot of my argument, then, is that we need to keep very clear the moral distinction between what the patient did choose and what the patient would choose. If there is insufficient evidence of actual patient choice, self-determination is irrelevant to the moral force of the “What would the patient choose?” question. That question should be restricted to cases without actual choice. In those cases, authenticity is relevant and important. Sometimes, however, authenticity will involve false beliefs, poor reasoning, and so forth. Then the right thing to do might be to accede to best interest considerations—that is, to accede to a view of the patient’s best interests, taken as a whole.

The point here is that although authenticity is highly important, it isn’t a trump. This raises a different and difficult question. When self-determination is invoked in clinical contexts, it is usually thought to override all other values.orage, if, by contrast, authenticity is not a trump, then in some cases other values might override it. Most of the time, it will be the patient’s own best interests that compete with authenticity, but other interests, such as the family’s interests, might also be relevant. (Of course, if we factor in the family’s interests, we will need some form of moral filter to keep the family’s financial interests from exploiting the patient.) And if the family’s interests are relevant, then sometimes those interests might tip the scale. Understanding the moral underpinning of the “What would the patient choose?” question as authenticity, not self-determination, makes such an outcome more likely, since authenticity seems easier to override than self-determination.

The decision-making might be further complicated if we distinguish between cases of permanent and temporary incompetence. The underlying analysis will be the same, but if we believe a patient is only temporarily incompetent, we might wonder whether “what the patient would choose” is entitled to greater weight because he will eventually be aware of the decision being made for him. The patient’s interests include not only living in accordance with his own values but being conscious of doing so. Suppose he would subsequently chafe at having had a choice made for him in opposition to his values. That itself would make his life worse.20

In the end, accepting my tripartite analysis might make the bedside decision quite a complicated one that will sometimes involve balancing values. If, for pragmatic reasons, simplicity is preferable, we could continue to treat the “What would the patient choose?” question as a trump—that is, we could overlook the conceptual mistake in the standard account of that question’s moral force. I think this would be a bad idea, but that is a topic for a different paper. Here, my concern has been simply to distinguish the different moral bases of the different questions a surrogate might have to answer.

Acknowledgments

I would like to thank Farr Curlin, John Lantos, and Lainie Ross for their very helpful comments on earlier versions of this paper.

References

1. See endnote 15.


7. To see authenticity as a value is not to be committed to the suspect claim that there is a “true me.” I use authenticity as a normative, not a metaphysical category. It involves only the claim that living in accordance with one’s beliefs and values is (usually) a good thing. Whether my beliefs and values express the true me is a separate and, here, irrelevant issue. I thank Farr Curlin for pressing me to note this issue.

8. For a powerful argument that the policy of making advance directives is, at least in practice, something that should be changed, see A. Fagerlin and C.E. Schneider, "Enough: The Failure of the Living Will," Hastings Center Report 34, no. 2 (2004): 30–42.


11. An excellent presentation of this distinction is in Buchanan and Brock, Deciding for Others, 116.

12. For the view that the moral justification for substituted judgment is that it ad-

13. This is the case, for instance, with Hobbes’s and Locke’s accounts of the social contract.


15. It is hardly definitive evidence, but it may be of interest to present the first two hits that came up when I Googled “substi- tuted judgment.” Consider a definition of “Substituted Judgement” (sic) from Ascension Health, found at http://www.ascensionhealth.org/ethics/public/issues/substituted.asp: “The form of surrogate decision- making regarding end-of-life care in which the surrogate attempts to establish with as much accuracy as possible what decision the patient would have made if that patient were competent to do so. This conclusion can be based on the patient’s preference expressed in previous statements or the surrogate’s knowledge of the patient’s values, beliefs, personality, and prior life style. This standard seeks to preserve the patient’s right of self-determination by placing the patient’s own preferences at the center of deliberation, while recognizing that it is the exception rather than the rule that the patient will have articulated his or her preferences in advance. Substituted judgment is to be distinguished from a decision based on an advance directive.” This definition conflates the issue of what the patient did say (“previous statements”) with the issue of evidence of what she would say (“the surrogate’s knowledge of the patient’s values, beliefs, personality, and prior life style”).

The second Google hit gave the following opening sentence of an abstract from Annals of Internal Medicine: “Substituted judgment has been proposed as a method of promoting the autonomy of the mentally incapacitated patient.” So far as “autonomy” includes self-determination, a mentally incapacitated patient has no such thing to promote. See A.B. Seckler et al., “Substituted Judgment: How Accurate Are Proxy Predictions?” Annals of Internal Medicine 115, no. 9 (1991): 743-45.

16. Even the best have faltered here, including Beauchamp and Childress and Buchanan and Brock. Both pairs of authors insist that substituted judgment should not be used for beings that have never been autonomous—for instance, someone who has been profoundly mentally disabled since birth—because such individuals were never autonomous and so cannot be autonomous-by-proxy. In fact, no one can be autonomous-by-proxy, if “autonomous” means self-determining. What these authors might be trying to get at is that the “What would the patient choose?” question can be relevant only if the patient was, at some point, capable of having a set of beliefs and values that were sufficiently rich that it makes sense to think of her life being or fail- ing to be in accordance with them. But the relevance to such a patient of the “What would the patient choose?” question cannot be a vindication of her (now nonexistent) capacity for self-determination. Both pairs of authors verge on making the point I am after, but neither quite does so. See T.L. Beauchamp and J.F. Childress, Principles of Biomedical Ethics, fifth edition (Oxford, U.K.: Oxford University Press, 2001), 99-100, and Buchanan and Brock, Deciding for Others, 112.

17. I thank John Lantos for pressing me to deal with cases of this kind.


20. I thank Lainie Ross for urging me to consider this additional complexity.