Confusion over Carnot's Theorem
PHYS 3350, D. Clarke, Jan. 2016

What's become known as *Carnot's Theorem* came from Sadi Carnot's book *Reflections on the Motive Power of Fire* first published in 1824, rather forgotten, then uncovered and published again with a translation in 1897. *Motive power* was an early term for *work* which, as interpreted by Güémez *et al.* (2002, Am. J. Phys., vol. 70, p. 42), Carnot defined as:

> [...] *the useful effect that an engine is capable of producing. The effect can always be expressed in terms of a weight being raised to a certain height.*

Using various *working substances* (gases, liquids) run repetitively through (reversible) thermodynamical cycles, Carnot concluded the following (also interpreted by Güémez *et al.*):

1. *Wherever there is a difference in temperature* [between the heat reservoirs], *motive power can be produced.*
2. [...] *the maximum amount of power gained by the use of steam* [in a Carnot cycle] *is also the maximum that can be obtained by any means whatsoever.*
3. *The motive power of heat is independent of the working substances that are used to develop it. The quantity is determined exclusively by the temperature of the bodies between which, at the end of the process, the passage of caloric*[1] *has taken place.*

From these statements, the literature (including undergraduate texts) have formulated *Carnot's Theorem* in predominantly one of two ways:

A. *The efficiency of all ideal (reversible) heat engines is the same, depending only on the temperatures of the heat baths from which heat is extracted and to which heat is ejected*;

B. *The efficiency of the Carnot cycle is greater than all other reversible or irreversible thermodynamical cycles, and is independent of the working substance used to convert heat into work.*

While statements like A claiming all reversible engines are equally efficient can be found in perhaps 3 out of 4 items in the literature (including text books and on-line physics resources), it is incorrect; B is the more accurate representation of Carnot's theorem. So how did interpretation A gain such a foothold?

The wording of Carnot's statement 2 is perhaps unfortunate. Without the words of clarity added by Güémez *et al.* (namely "in a Carnot cycle" justified from earlier portions of this statement represented by [...]), statement 2 could be interpreted to mean that any reversible

---

[1]In 1789, Antoine Lavoisier coined the term *caloric* to refer to a weightless gas contained by all matter that carried heat from one object to another when caloric in the two objects was unbalanced. The idea was replaced by the first law of thermodynamics when it was introduced in the mid 19^{th} century.

thermodynamical cycle whatsoever has the same efficiency. In fact, what Carnot meant was that specifically the Carnot cycle, using any *working substance* whatsoever, has the same efficiency. In other words, the efficiency is independent of the *substance*, not the *cycle*.

Those inclined to take the incorrect interpretation of Carnot's second statement might then find solace in the third. The last sentence, for example, taken out of context could be construed to claim that work extracted from *any* reversible cycle depends only on the temperatures in the heat reservoirs, and thus the same as the Carnot cycle. However, in the sentence immediately preceding, Carnot clearly states that he is talking about independence from the working substance, not necessarily thermodynamical cycle.

In his text *Classical and Statistical Thermodynamics*, Ashley Carter is among those adopting interpretation A and thus propagating this confusion. His §6.4 concludes on page 94 with:

> *That is to say,* all [his emphasis] *reversible engines operating between the same reservoirs have the same efficiency,* $\eta = 1 - T_{\mathrm{L}}/T_{\mathrm{H}}$. *Irreversible engines will have a lesser efficiency. This is Carnot's Theorem.*

To be clear, this is patently false.

Let's examine his "proof" of this conclusion—one you'll find in numerous texts as well as Wikipedia, Khan's Academy, *etc.*—which is predicated on his Figs. 6.7, 6.8, and 6.9. In each figure, notice that two very important assumptions are made without mention, and then unknowingly—I have to assume—transmitted into the mathematics:

1. *All* heat enters and exits the system from constant-temperature reservoirs and work is then computed as a difference of these heats. Thus, these portions of the thermodynamical cycle must be isotherms.

2. With all heat passing through the isotherms, there is none left to enter or exit from the subprocesses that join the two isotherms (which must exist lest the two isotherms be the same). These, therefore, must be adiabats.

So, how many thermodynamical cycles are composed of two isotherms and two adiabats? **Just one: the Carnot cycle!** So, what Carter actually proves is:

> All reversible *Carnot-cycle* engines operating between the same reservoirs have the same efficiency,

which is much like saying 'All red apples are red.' Nothing to argue with, just not particularly profound!

Carter is not alone. Google "Carnot's theorem", and you'll find numerous examples of professional physicists asserting that Carnot's theorem states that all reversible cycles have the same efficiency as the Carnot cycle. It just isn't true!

Indeed, when the third edition of the fabled first-year text *Physics (for students of science and engineering)* by Halliday and Resnick became Halliday, Resnick, and Walker in edition

4, the example of a Stirling engine was introduced in their Chapter 22 on entropy. Their Figure 22-5 (which became their Figure 21-8 in the fifth edition and reproduced in the left panel of the figure below) shows the Stirling cycle with two isotherms and two *isochors* with heat entering and leaving the system *only* through the isotherms. They also offered a "proof" that this and any reversible engine has the same efficiency as the Carnot cycle which, while quite a bit simpler than Carter's "proof", is equally wrong. They start off correctly by stating that entropy, being a state variable, doesn't change in a complete thermodynamical cycle. Thus,

$$0 = \Delta S = \oint \frac{\text{đ}Q}{T} = \frac{Q_{\rm in}}{T_{\rm H}} + \frac{Q_{\rm out}}{T_{\rm L}} = \frac{Q_{\rm in}}{T_{\rm H}} - \frac{|Q_{\rm out}|}{T_{\rm L}}$$

$$\Rightarrow \quad \frac{Q_{\rm in}}{T_{\rm H}} = \frac{|Q_{\rm out}|}{T_{\rm L}} \quad \Rightarrow \quad \frac{|Q_{\rm out}|}{Q_{\rm in}} = \frac{T_{\rm L}}{T_{\rm H}} \quad \Rightarrow \quad \eta = 1 - \frac{|Q_{\rm out}|}{Q_{\rm in}} = 1 - \frac{T_{\rm L}}{T_{\rm H}}.$$

They continue by concluding that since the ratio of the temperatures does not depend upon the nature of the reversible cycle, the efficiencies of all ideal (reversible) engines must be the same, as is stated clearly in their equation 21-9 in edition 5.

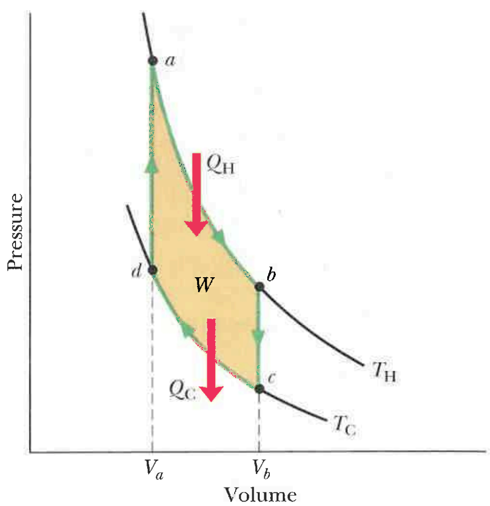


FIGURE 21-8 A p-V plot for a Stirling engine. The isothermal expansion *ab*, the isothermal compression *cd*, and the two constant-volume processes *bc* and *da* correspond to the four strokes of Fig. 21-7.

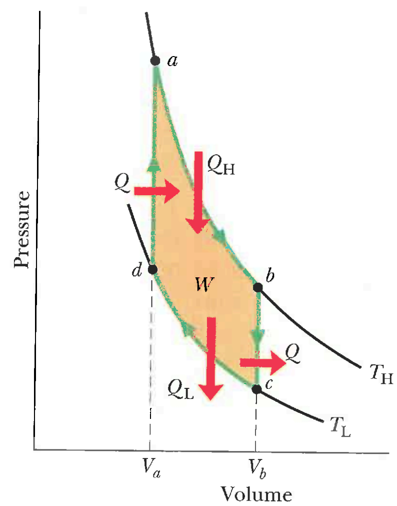


Fig. 21-12 A p-V plot for the working substance of an ideal Stirling engine, assumed for convenience to be an ideal gas.

Can you spot the flaw in their argument? In order to perform the integral as they did, they had to assume that where $\text{đ}Q \neq 0$, T is constant. Thus, just like in Carter's argument, they assume implicitly that *all* heat enters and leaves through isotherms, leaving no heat to enter or leave the system through the adjoining sub-processes. These, therefore, must be adiabats, once again identifying the cycle uniquely as a Carnot cycle, not a Stirling cycle.

It is telling (and reassuring) therefore, that this was corrected in their edition 6 (*e.g.*, their Figure 21-12 reproduced above right). Here, notice the addition of the same Q entering the isochor on the left and leaving the isochor on the right. (They must be the same because the heat entering/leaving an isochor depends only on the temperature difference of the states they connect which is the same given that the isochors are joined by isotherms.) One might protest that if the same amount of heat entering one isochor leaves the other, surely these effects must cancel!

Not at all. In this case, the efficiency is given by:

$$\eta_{\text{Stirling}} = 1 - \frac{|Q_{\text{out}}|}{Q_{\text{in}}} = 1 - \frac{|Q_{\text{L}}| + |Q|}{|Q_{\text{H}}| + |Q|},$$

where $|Q_{\text{H}}|$ and $|Q_{\text{L}}|$ are the quantities of heat entering and leaving the system along the high- and low-temperature isotherms respectively, and $|Q|$ is the quantity of heat entering and leaving the system along the isochors. It should be immediately apparent, then, that η falls as Q increases.

More specifically, along the isotherms and referring to HRW's Fig. 21-12 above, we can write:

$$Q_{\text{H}} = RT_{\text{H}} \ln(v_b/v_a); \qquad Q_{\text{L}} = -RT_{\text{L}} \ln(v_b/v_a);$$

whereas along the isochors, we have:

$$|Q| = c_v(T_{\text{H}} - T_{\text{L}}).$$

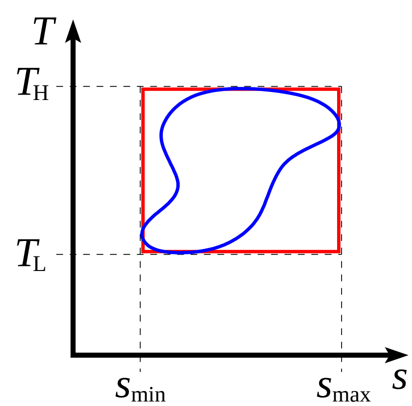
Thus,

$$\eta_{\text{Stirling}} = 1 - \frac{RT_{\text{L}} \ln(v_b/v_a) + c_v(T_{\text{H}} - T_{\text{L}})}{RT_{\text{H}} \ln(v_b/v_a) + c_v(T_{\text{H}} - T_{\text{L}})} = 1 - \frac{T_{\text{L}} + T_{\text{off}}}{T_{\text{H}} + T_{\text{off}}},$$

where,

$$T_{\text{off}} \equiv \frac{c_v}{R} \frac{T_{\text{H}} - T_{\text{L}}}{\ln(v_b/v_a)},$$

is the effective *offset temperature*, which renders the efficiency of the ideal Stirling engine less than that of the Carnot cycle.



By the way, the proof of interpretation B is actually quite simple. Construct any loop on a T-s diagram bound by an entropy domain, $s_{\text{min}} \leq s \leq s_{\text{max}}$, and a temperature range, $T_{\text{L}} \leq T \leq T_{\text{H}}$ (*e.g.*, the blue loop in the figure). On a T-s diagram, the area under the upper/lower bound of the loop is the heat entering/leaving the system, while the area within the loop is the work done per cycle. One can then see by inspection that to maximise the heat entering the system per cycle, Q_{in}, and at the same time minimise the heat ejected, $|Q_{\text{out}}|$, the loop must be a rectangle (red) and thus bound by two isotherms and two adiabats. This is precisely the Carnot cycle. Any other loop bound by $s_{\text{min}} \leq s \leq s_{\text{max}}$ and $T_{\text{L}} \leq T \leq T_{\text{H}}$ necessarily has a higher value of $|Q_{\text{out}}|/Q_{\text{in}}$, and thus a lower efficiency[2]. Note that this argument has no dependence on the nature of the working substance undergoing the thermodynamical cycle. So long as the substance can physically attain each state around the closed loop, the efficiency of the cycle depends only upon the geometry of the loop.

[2] And with a little bit of thought, one can eliminate the need for the loop to be bound in entropy, leaving the bound in temperature the only relevant one.